EVENT DETECTION FOR POST LUNG TRANSPLANT BASED ON HOME MONITORING OF SPIROMETRY AND SYMPTOMS

A DISSERTATION

SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL

OF THE UNIVERSITY OF MINNESOTA

BY

XUEWEI WANG

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF

DOCTOR OF PHILOSOPY

STANLEY M. FINKELSTEIN, PH. D.,
ADVISOR

DECEMBER, 2011

ACKNOWLEDGMENTS

This research was supported in part by NIH grants R01NR02128 and R01NR009212.

I wish to thank professor Stanley M. Finkelstein, my doctoral advisor, for his guidance, thoughtful inputs and patience over the years supervising my research. I would like to thank my committee members: professor Ahmed Tewfik, for his work and direction that set the theoretical foundation of this research; professor William Thomas, for his meticulous review and comments which greatly helped me in the organization and writing of this dissertation; professor Robert P. Patterson, for his insightful suggestions and mind stimulating discussions; and professor James Holte for his enthusiasm in the topic and accommodation to committee changes. I also would like to thank professor Marshall I. Hertz, MD, for his help in gaining clinical insights as well as in manuscript reviews.

I thank Jeffery Lande, John Troiani, Kathleen Harrington and Virgil Larson for contributing to the research material.

I thank my colleagues at work, Qingshan Ye, Zengri Wang, Jian Cao, Leonard Leuer for stepping out of their ways and help review my writings.

Finally, I thank my wife Xin and my adorable daughter Meryl for their love and support over the years.

DEDICATION

This work is dedicated to my parents

此文敬献我的父母

ABSTRACT

The goal of this dissertation research was to develop, implement, and test an automated decision system to provide early detection of actual acute bronchopulmonary events in a population of lung transplant recipients following a home monitoring protocol.

Decision rules were developed using wavelet analysis of spirometry and symptom signal data collected daily at home by the lung transplant recipients, and transmitted weekly to our study data center. Rules were developed based on a learning set of patient home data, and validated with an independent set of patients.

Using either FEV_1 or symptom-based home data monitoring, the detection algorithm can capture the majority of events (sensitivity > 80%) at an acceptable level of false alarms. Detection occurs 6.6 to 10.8 days earlier than the corresponding events recorded in the patient's clinical records. Combining rules using the Dempster-Shafer theory of evidence incrementally improves performance over a single variable.

This framework can be readily implemented as an automatic event detection tool to aid medical discovery and diagnosis of acute pulmonary events.

Keywords:

Lung transplant, home monitoring, spirometry, pulmonary function testing, event detection, wavelets, CUSUM, operating characteristics, Dempster-Shafer, classifier combination

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

KEY ABBRIEVATIONS

LTHMP Lung transplant home monitoring program at the University of Minnesota

ISHLT International Society for Heart and Lung Transplantation

LTx Lung transplant

BO/BOS Bronchiolitis Obliterans/ Bronchiolitis Obliterans Syndrome

PFT Pulmonary function test

FEV_1 Forced expiratory volume in 1 second; volume of air forcibly expired from a maximum inspiratory effort in the first second (L)

FVC Forced vital capacity; the total volume that can be forcefully expired from a maximum inspiratory effort (L)

FFr FEV_1 to FVC ratio

FEF Forced expiratory flow.

PEFR/PFR Peak expiratory flow rate; the highest forced expiratory flow (L/second)

MEFR/ FEF_{25-75}/MFR Mid expiratory flow; the forced expiratory mid flow rate at 25 to 75 percent of forced vital capacity (L/second)

SNR Signal to noise ratio

MA Moving average

WMSA Wavelet-based multiscale analysis

DWT Discrete wavelet transform

OC Operating Characteristics

CUSUM Cumulative sum

DST Dempster-Shafer (DS) theory of evidence

PDF Probability distribution function

bpa Basic probability assignment

BOE Body of evidence

Chapter I

INTRODUCTION AND OVERVIEW

Introduction

Lung transplant recipients are at high risk of post surgery complications such as infection and rejection, two leading contributors to morbidity and mortality [1]. Patient survival remains inferior to other organ transplantations such as heart and liver [2 , 3 , 4 , 5], primarily due to severe consequence of allograft-related complications. Many transplant centers recommend that lung recipients perform daily monitoring at home, which may detect early signs of transplant complications and thus enable timely diagnosis and treatments [6,7,8]. Survival of lung transplant recipients is enhanced when infection or rejection is detected early and appropriate therapies are implemented [9,10]. Early intervention may also reduce cost and improve quality of life [11]. Therefore, early detection of lung transplant complication events is of paramount importance to these patients.

Studies have shown that it is feasible to detect general clinical problems early by monitoring pulmonary functions (spirometry) and symptoms from patient's home [12, 13]. Spirometry quantifies pulmonary function changes by measuring the amount and speed of air that can be inhaled and exhaled with maximal effort in limited time, therefore, provides mechanistic insight into the physiologic impairment of the lungs. Spirometry is known for recognizing pulmonary pathology [14] and is routinely used to aid diagnosis of lung disease [15]. Reliable and valid spirometry measurements can be obtained by patients in a home setting and are comparable to the measurements obtained in clinic [16]. Home based testing with a portable spirometer permits higher testing frequency as compared to in-clinic visits, therefore, enables early detection of clinical problems. In particular, in the lung transplant home monitoring program at the University of Minnesota (LTHMP),

Finkelstein and colleagues demonstrated that home spirometry could detect Bronchiolitis Obliterans Syndrome staging, a syndrome considered to be related to chronic rejection [12] at its early stage. An automated triage decision support system can identify which participants have demonstrated critical level of concerns to be on a closely “watched” list [17]. At other institutions, Wagner and colleagues reported telemetric monitoring for reliable early diagnosis and treatment of infection and rejection [14]. Morlion and colleagues reported an internet-based home monitoring approach to detect acute events [18]. The possibility of early detection of related pathologies through home monitoring has thus been fairly well documented for the bronchopulmonary conditions.

Challenges and Motivations

Recent technological advancements have made home monitoring more convenient and more reliable. However, new challenges emerge as more data need timely processing and relevant decisions. From the clinician’s perspective, one key objective that remains unresolved is how to utilize the home data to detect clinical events in a prompt and reliable manner.

The long term home monitoring for the lung transplant recipients in a clinical setting presents numerous challenges: Data are extremely precious and there are no public data available. There is paucity in public research over the last decade per a Medline/PubMed database search using the key words of “spirometry monitoring”. To this end, this research studied proprietary data from two cohorts of consented lung transplant recipients in LTHMP, and investigated methods customized for this type of data.

As to be illustrated, the home spirometry data series are nonlinear nonstationary with wide range of inter- and intra- subject variability, which can be difficult for clinicians to interpret. It is unknown if the data fluctuation encodes pertinent event information, and if it does, how an event will exhibit itself in data and how best to decode it. The underlying disease development and progression can be masked among other factors such as inconsistency in patient’s physical

effort. The data variation due to an event needs to be separated from the background noise. The problem can multiply as the types of events and patients' reaction to events vary. Therefore, simple guidelines and heuristics do not readily meet the challenges of the complexity in data for accurate signal estimation and subsequent event detection.

Another challenge involves how to retrieve relevant information from the different data types and reach reliable decisions quicker. This problem is two-fold:

On one hand, clinicians have busy schedules. Modern home monitoring technology often provides great amount of data, which would not be useful if they can not be timely digested and utilized. In LTHMP, patients were instructed to perform daily spirometry and record symptoms along with vital information. A home monitoring report may consist of up to ten spirometric, five symptom, five health status/ well-being, and five vital sign variables. These multidimensional data accumulate over time and increase the burden to data review. One must determine which monitoring variables are most essential.

On the other hand, event detection based on a single variable input can be suboptimal. For example, Morlion's study reported a sensitivity of 63% in detecting acute rejection based on spirometry monitoring and questioned the potentially missed detection by spirometry alone [18]. Therefore, there is a desire to combine different classes of monitoring data for better detection. To date, there is paucity in this domain literature in regard to how to meaningfully combine such information. This prompts investigating techniques to boost detection performance. However, distinctive types of variables are interpreted very differently. Each data type requires customized processing. This is because spirometry and symptoms acquire their evidence based on totally different criteria: spirometry is effort-based and equipment measured, and symptoms are self assessed. Generally, such distinctive data types can not be combined due to different characteristics and embedded noise. Therefore, it is a technical challenge to combine the spirometry and symptom information.

Significance

The critical nature of event detection suggests that effective and robust means be deployed to make use of suitable monitoring information.

The significance of the research resides in the following:

Contribution to lung transplant home monitoring. This research explores new signal analysis methods to detect events. Signal analysis is the least invasive approach, is compatible with the existing home monitoring infrastructure, and can be readily put into use. A systematic data screening approach could have a profound impact to the post-lung transplant surveillance and care. The resulting methods would provide automatic just-in-time detections. By establishing a well understood detection mechanism with clinically meaningful performance, this research would demonstrate further evidence of clinical utility, and might increase the acceptance of home monitoring of lung transplant patients.

Patient welfare. Effective non-invasive computerized event detection should be pertinent and cost-effective for the detection of major transplant problems. Acting upon timely diagnosis, cost effectiveness, patients' quality of life and survival are expected to improve.

Research enrichment. "Quantifying and modeling the complexity of health variability, and detecting more subtle alterations with disease and aging, present major challenges in contemporary biomedicine" [19]. Although the focus is on one discipline (lung transplant home monitoring), this work has general implication to the research of the long term care of the chronically ill.

Objectives and Specific Aims

Previously, a decision support nursing triage algorithm has been developed to detect *potential* bronchopulmonary problems [17] for the LTHMP. This research aims to step further and develop a method that automatically detects *actual* clinical events based on the same home monitored spirometry and symptom data. The events are the actual events captured in the patient's medical records.

The first objective is to understand and characterize the home monitoring data and to determine an effective analytical approach. Secondly, it was a design factor that the detection could be implemented in a non-invasive computerized framework which eventually will operate in real time. The framework shall work within the existing infrastructure of data collection and assimilation of LTHMP.

Previous research approaches [6,12,14,17,18, 20] have been critically reviewed. As the home monitoring data are noisy, nonlinear and nonstationary, this research first proposed a wavelet-based method which is a natural tool for multiresolution time dependent analysis and is well known to handle nonlinear nonstationary data. Wavelet method is the analytical foundation for event detection in this study.

Next, based on the output of individual classifiers, a multi-classifier combination was implemented to enhance detection over the univariate approach. This combination rule is based on the Dempster-Shafer theory of evidence (DST).

The desired algorithms are data driven, individually customizable and operable in real time. These algorithms can be used as decision-support to detect clinical events and to aid medical discovery/diagnostics. The ultimate goal is to improve lung transplant patient survivability. This decision aid is complementary and not a replacement to the primary care [21].

The specific aims are to:

1) Demonstrate the utility of wavelet-based method for characterizing and analyzing home monitored data.

2) Derive event detection features to be tested with actual events.

3) Demonstrate clinically relevant performance metrics which include sensitivity, specificity/ false alarm rate and timing of detections.

4) Validate the detection design with broader previously unseen patient set and demonstrate reproducible results. The results from this research may possibly serve as a benchmark for future research with similar aims.

This research consists of a two-stage structure of event detection, outlined in Figure 1. The first stage involves individual event classifier development based on the wavelet analysis. The second stage involves a multiple classifier combination based on the Dempster-Shafer theory. Performance is evaluated in the same manner for each approach.

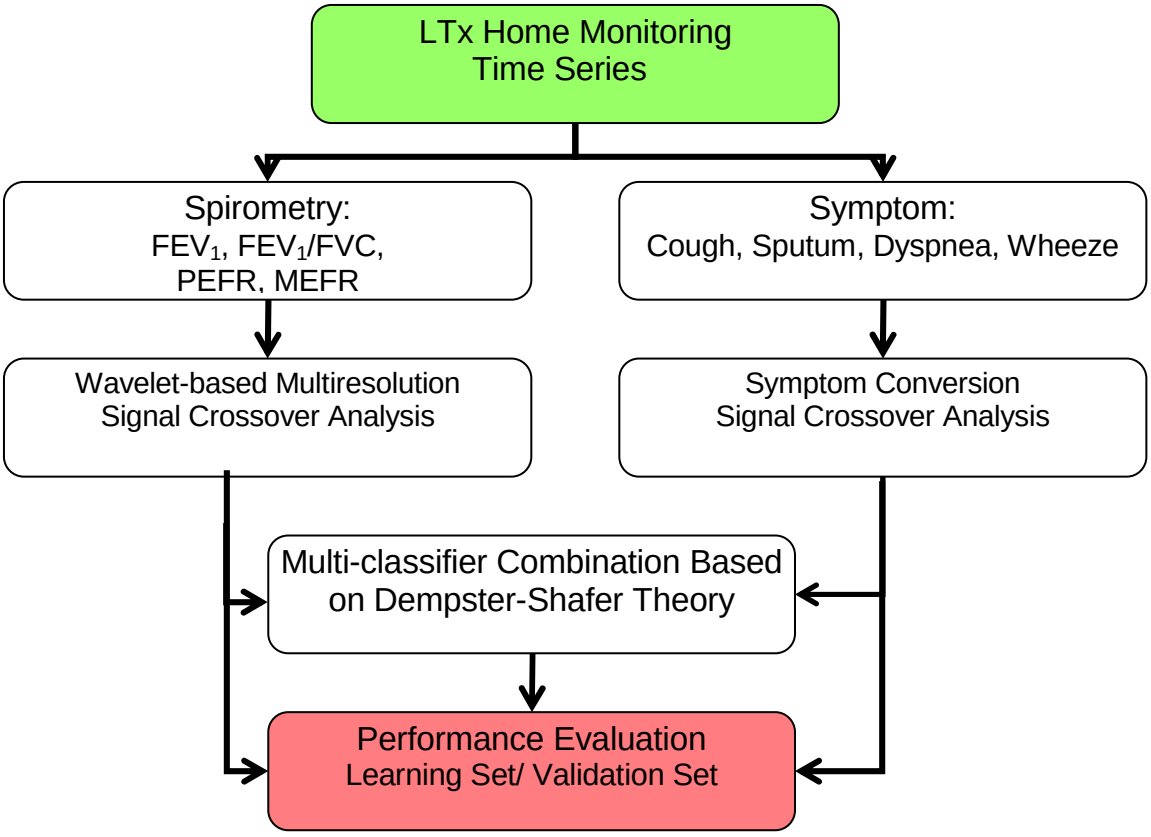

Figure 1. Topics of the dissertation research

Note: Abbreviations are noted on page vii.

Organization of the Dissertation

Chapter I provided introduction and overview, challenges and motivations, objectives and specific aims, significance and scope of the research. The rest of the dissertation is organized as follows:

Chapter II describes domain specific background information and a literature review in topic areas concerning this research. The literature review also provides rationale and background of the chosen approaches in this research.

Chapter III describes the materials that were used in this research. In particular, these datasets were used in developing, learning and validating the methods. Data characteristics are illustrated and summarized. Pre-analysis data processing was performed.

A two-staged approach to event detection was implemented:

Chapter IV focuses on the wavelet approach to data analysis, processing and derivation of univariate classifier, with the theoretical background outlined first. Based on a wavelet-estimated signal crossover, a sequential cumulative sum (CUSUM) classifier is proposed for event detection. Event detection and classification are illustrated through cases. Performance metrics are defined. Performance is evaluated for the learning set and the validation set, respectively.

Chapter V focuses on the Dempster-Shafer approach to multi-classifier combination. Parallel to the wavelet-based univariate classifier, methods are illustrated with cases, and overall performance is evaluated for the learning set and the validation set, respectively, followed by comparing to results in Chapter IV.

Chapter VI discusses the clinical relevance, impacting factors, utilities and limitations of the research.

Chapter VII summarizes the contribution to research, concludes, then suggests future work.

Appendices consist of additional technical background, the supporting work, documentation of computer code, and related work not included in the main body of the dissertation.

Chapter II

BACKGROUND AND LITERATURE REVIEW

Lung Transplant

Lung transplantation has become an acceptable therapeutic option of many end stage lung diseases [2,22].

Since the inauguration of the annual report of the Registry of the International Society for Heart and Lung Transplantation (ISHLT) in 1984, there is a steady increase in the cases reported over the years as more institutions and databases worldwide are linked each year to report the prevalence [23,24]. Approximately 30,000 adult lung recipients and 3500 adult heart-lung recipients have been reported around the world through 2008 in the ISHLT 2009 annual report [4]. In the US, the prevalence of people living with a functioning lung transplant at end of the year was 2,099 in 1995, 3,400 in 1999, 4,853 in 2003 and 6,731 in 2007 [25, 26, 27, Figure 2]; For instance, there were 1,085, 1,468 and 1770 lung transplants performed in 2003, 2007 and 2010, respectively [28,29].

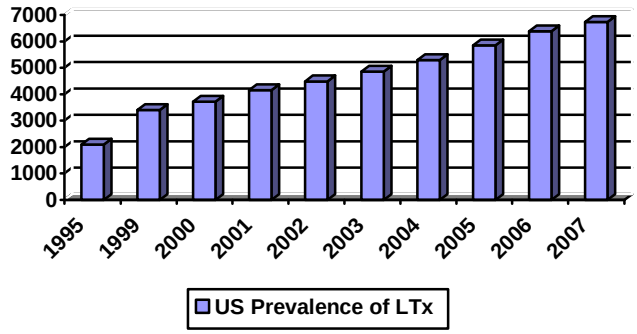

Figure 2. Prevalence of lung transplant in the US.

The need for lung transplantation is due to lung failure. The most common diseases that result in lung failure include chronic obstructive pulmonary disease, emphysema, cystic fibrosis, idiopathic pulmonary fibrosis, primary pulmonary hypertension, bronchiectasis, sarcoidosis. The three main types of lung transplant are single lung, double/bilateral lung and heart-lung transplant [30].

Certain physiologic changes are expected after lung transplantation. In particular, two important host defenses to lung infections are impaired: cough reflex and mucociliary transport. Since patients don't necessarily have sensation of the new denervated organs [31] and may have no early pain symptoms of a blockage building up, monitoring cannot rely on symptoms alone, especially in the early stage of 4-6 weeks post transplant. Pulmonary function should rise and level off during the first 3 months post transplant. Any time after the first 3 months, a sustained decline of pulmonary function could represent a potential problem [10].

The common adverse effects after a successful transplant are rejection (acute and chronic) and infections. After a transplant, it is necessary to take immunosuppressants or antirejection medications to suspend the body's immune system and protect the transplant. The medication reduces the risk of organ rejection, however, it also increases the risk of infections in the newly transplanted organ. In lung transplantation, the organ is not matched to the patient except by a blood type, thus the transplant tissue could be very different from the native tissue. This means that rejection is more common and is often more severe than in other organ transplants such as kidney transplantation. The chronic rejection is also denoted as bronchiolitis obliterans (BO), in which there are obstructive changes in the airways of the lungs. Other effects include high blood pressure, diabetes mellitus, high cholesterol, reappearance of lung disease, cancer, osteoporosis, cataract, and kidney disease. Nearly all patients will have at least one or more rejection episodes after transplant surgery [32].

Infections are more likely since there is direct exposure to pathogens in the environment [33]. Community acquired respiratory viruses and bacteria, such as

Methicillin-resistant Staphylococcus aureus (MRSA), has also been recognized as a major cause of morbidity and mortality [34]. Graft failure and non-cytomegalovirus (CMV) infections were the principal fatal complications in the first 30 days, and bronchiolitis obliterans/ Bronchiolitis Obliterans Syndrome (BOS), a manifestation of chronic rejection, has been the single largest contributor to late mortality [2,23]. About 50% of lung recipients [23], and 30% to 50% percent heart-lung recipients develop bronchiolitis obliterans post-transplant [32]. Once established, bronchiolitis obliterans is difficult to treat.

Acute rejection occurs in 35% to 50% of lung transplant recipients and is an important risk factor for chronic rejection/ bronchiolitis obliterans. Acute rejection, by itself, is usually not life or graft threatening, and its abnormal histology can be reversed with treatment [35]. Thus, early diagnosis and treatment with augmented immunosuppression might help reduce the amount of irreversible parenchymal injury in the allograft [14] , and offer the best hope of decreasing the lethality of this complication [2, 36 , 37]. Rejection is characterized by fever, dyspnea, cough, sputum production, rapidly developing pulmonary infiltrate, and pulmonary decompensation [36,38] such as decreased SaO_2 (oxygen saturation test) and forced expiratory volume at one second (FEV_1).

To carefully monitor transplant patients for signs of rejection, small pieces of the transplanted organ are removed for inspection under a microscope. This is called a biopsy. Transbronchial biopsy (TBB) is a non-surgical procedure performed using a bronchoscope and special biopsy forceps designed to be used in the lung. During a transbronchial biopsy, biopsy forceps are used to retrieve a small piece of tissue from a suspicious area of the lung or airway. The tissue sample is then examined to diagnose lung disease. Bronchoalveolar lavage (BAL) is a medical procedure in which a bronchoscope is passed through the mouth or nose into the lungs and fluid is squirted into a small part of the lung and then recollected for examination. Transbronchial biopsy and bronchoalveolar lavage are often used lung biopsy procedures. The sensitivity of TBB for diagnosis of

bronchiolitis obliterans (BO) varies from 15 to 78%, and the specificity varies from 75 to 93% in various studies [39]. BAL has not been proven useful in the diagnosis of chronic rejection (BO). Lung biopsy may be considered "gold standard" of medical diagnosis of rejection, BO and infection, but it is invasive, uncomfortable, costly, and often has its own associated morbidity and mortality [1]. Primary graft dysfunction frequently complicates lung transplant and has been an area of intense study [34]. The search for molecular biomarkers of acute and chronic lung rejection is underway. Currently biological diagnosis based on microarray analysis uses gene expression to predict and characterize acute and chronic rejection. However, the application of new genetic, genomic, and proteomic technologies is in its infancy for application to lung transplantation [35].

Noninvasive surrogates to biopsy include exhaled nitric oxide, computed tomography (CT), and pulmonary function testing/ spirometry [40]. Chest CT may not always be effective in detecting rejection [32] since the interstitial infiltrate seen on x-ray is hard to distinguish rejection from that of an infection. Spirometry provides measurement of the lung function but lacks the specifics to determine etiology [2].

Telemedicine and Home Monitoring

Technological advancement has enabled patient care beyond the boundary of hospital walls. In LTHMP, data are measured by patients at their homes, then transmitted, and analyzed at the data center. This new type of patient care, referred as telemedicine, concerns performing diagnostics and delivering treatment to patient remotely away from the clinic using computers and telecommunication technologies [41, 42]. The idea behind telemedicine is to provide more convenient and customized care to patients, using such technologies as video conferencing, smart phones, and other wireless devices to interact with patients in their homes. Telemedicine can provide more comprehensive and monitored health care at home such as remote monitoring of pulmonary function, blood pressure, ECG, glucose, vital data [43] and is also a logical option to reach

remote areas [44]. In many cases, patients are capable of performing the requested physiological measurements. Data communication infrastructure has advanced from the earlier telephone-modem, to digital communication systems including the internet and wireless communications that are adding convenience and driving down the cost of such a system. Data acquisition has been made easier to use with miniaturized and wearable devices, and less effort from the patients. Virtual home visit systems had been implemented [45,46,47] and the perceptions from patients were largely positive [48, 49]. In the post-lung transplant care, Internet-based home spirometry-monitoring [18] as well as other electronic monitoring programs [50] have been implemented. Telemedicine has moved beyond mere data transfer to the launch of novel applications like telesurgery, telerehabilitation, telecardiology, remote patient management and a whole spectrum of teleapplications [51, 52]. “Technology innovation in healthcare contributes to increased independence and greater empowerment of patients as participants in their own medical care [53].” As the health care system moves further toward a patient-centered, ambulatory, and home-based approach to care, the use of home monitoring will become both more prevalent and more necessary [17].

From the cost perspective, home monitoring may revert expensive inpatient care to less expensive outpatient care. It may reduce the expenses by avoiding unnecessary clinic or emergency visits and the high maintenance due to late diagnosis. A study of cost-effectiveness of LTHMP indicated that at one-year post transplant, the average cost saving was about $11,000, and patient survival was higher in the adherent patients than in the non-adherent patients [11]. Although no definitive work has been performed to demonstrate the improvement in quality of life via home monitoring, patients are being monitored in the comfort of their home setting.

Successful telemedicine implementations and induction into clinical routine requires well orchestrated effort. Technology, acceptance, financing, organization

and policy and legislation, often decide the inclusion of a telemedicine program into daily practice [54]. The new healthcare reform initiatives have made case for more research on means to reduce healthcare cost at same or enhanced quality. For LTHMP, the demonstration of clinical utility and outcome is key to persuade professionals, policy makers and insurance companies about the benefit of this program. Therefore, this dissertation research involves a small but key component in this broad picture.

Spirometry

The Latin word *spirare* means to breathe. Spirometry is a pulmonary function test (PFT) that describes the ability of the respiratory system to move air by timed respiratory volumes [55]. PFT, as described in American Standardization of Spirometry, is an effort dependent maneuver. A subject inhales with maximum effort, and then blows the air out as forceful and fast as possible. Spirometry contains mechanistic insights into the physiologic impairment of the lungs. Flow/volume spirometry is routinely used for assessing the type and severity of lung disease in clinic. Following lung transplant, prompt diagnosis and intervention depends on monitoring key changes in pulmonary function [56]. "Recognize the (air flow) pattern, and the diagnosis will follow" [15].

The mostly commonly used indices describing the respiratory cycle are as follows:

- The forced vital capacity (FVC). A spirometric maneuver begins with the patient inhaling as deeply as he or she can. Then the patient exhales as long and as forcefully as possible; the amount exhaled in this manner is the FVC.
- The forced expiratory volume in one second (FEV_1). FEV_1 is the amount of air exhaled during the first second of the FVC maneuver. It tends to be lower in diseases that obstruct the airway, such as asthma or emphysema.
- The FEV_1/FVC ratio (FFr) is used to determine if the pattern is obstructive, restrictive, or normal.

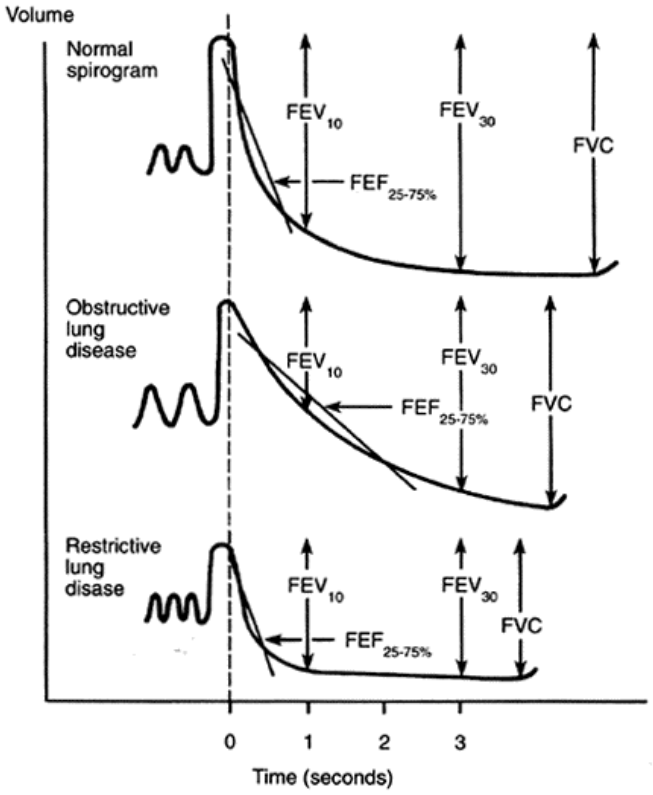

Figure 3. Spirogram recording from Forced Vital Capacity Maneuvers

(Source: Respiratory Physiology [57]).

Note: Abbreviations slightly differ from the ones referred in this paper.

FEV_{10} refers to the forced expiratory volume in one second from time zero (FEV_1). FEV_{30} refers to the forced expiratory volume in three seconds from time zero. $FEF_{25\text{-}75\%}$ is the same as mid flow (MEFR).

Figure 4. The spirometer/diary instrument.

A spirometer/diary instrument (PFM-H100, Telemedical Inc., Minneapolis, MN, US) was designed for LTHMP to collect and store daily monitoring data. The device includes a portable spirometer along with a reusable, replaceable plastic mouthpiece, data input pad, digital display, internal modem, and printer [58].

The most widely used indices that are computed from the spirogram (Figure 3) include forced vital capacity (FVC), forced expiratory volume at 1 second (FEV_1), maximal mid expiratory flow rate (MEFR, also denoted as forced expiratory mid-flow rate or $FEFR_{25-75}$, the difference of flow rates at 25% and 75% of the FVC volume respectively), and peak expiratory flow rate (PEFR). Measures are often presented as a percent of predicted (normal) value. Predicted values of FEV_1 and FVC are based on population studies and vary by race, gender, age and height [59]. Spirometry features have been studied in identifying events [60]: FEV_1 is the most frequently used index for assessing airway caliber, airway obstruction, bronchoconstriction or bronchodilation and provides the best estimates of airflow obstruction in patients with asthma, chronic obstructive pulmonary disease (COPD) and emphysema. It is affected only partially by subject cooperation and is determined both by effort-dependent and effort-independent portions of the maximal expiratory maneuver and often with lower variability than other variables derived from the flow-volume curve [61]. FEV_1/FVC ratio (FFr) is another standard index for assessing and quantifying airflow limitation, and is also often used to determine if the pattern of spirometry is obstructive [60]. MEFR has been reported with good sensitivity for diagnosis of infection and rejection after heart-lung transplantation and is predictive of acute allograft dysfunction [62]. PEFR (peak

expiratory flow rate) is advocated and commonly used to monitoring patients with asthma [62,63]. A change in peak flow predominantly reflects alteration in the caliber of large airways as opposed to FEV_1 which is affected by change in caliber of both large- and medium-sized airways [64, 65]. Therefore, PEFR was considered indirect index of airway caliber and less sensitive than FEV_1 in detecting airway obstructions [61]. Spirometry measures can be indicative of pathological processes that can often be insidious while symptoms may not be apparent to recipients until the transplanted organ is severely compromised [58].

Table 1 provides definitions of a list of common spirometry variables.

Air flow obstruction is an early sign of pulmonary rejection and is correlated with declines in FVC, FEV_1, and MEFR [66]. Obstructive airflow is related to the narrowing of the airways due to bronchial smooth muscle contraction as in asthma; inflammation and swelling of bronchial mucosa and the hypertrophy and hyperplasia of bronchial glands as in bronchitis; material obstructing the flow of air as in excessive mucus plugging, foreign objects or tumors; destruction of lung tissue with the loss of elasticity and hence the loss of the external support [59].

Restrictive airflow patterns are associated with intrinsic disorder as in pneumonia and tuberculosis; extrinsic disorder as in tumors, pain on inspiration, and pleural effusion; neuromuscular as in paralysis of diaphragm. It cannot be determined by spirometry alone [59]. The total lung capacity must be assessed as well.

Clinic spirometry is routinely used to assess pulmonary function post lung transplant. Agreement between home spirometry and clinic spirometry was shown to be excellent [75]. Spirometry is used for early home recognition of asthma exacerbation in high risk patients with severe persistent disease, and for recognition of either infection or rejection in lung transplant patients [67]. It was pointed out that a fall in spirometry is seen in infection and rejection post-lung transplant and that the underlying disease state has a significant influence on the

diagnostic utility of specific spirometric indices [70]. Studies using TBB and BAL as gold standards have reported that the sensitivity of FEV_1 for the detection of infection/ rejection related complications ranges from 48% to 72% in recipients of single lung, and 60% to 75% in bilateral-lung and heart-lung transplantation [68,69,70]. These results have led to the widely accepted recommendation that the follow-up of lung transplant recipients should include daily measurement of FEV_1 at home with a portable spirometer [18].

In summary, studies showed acceptable performance by sensitivity/ specificity in the sequence of lung biopsy, clinic spirometry and home spirometry, each validated against its predecessor. Per ISHLT guideline, "FEV_1 is the most reliable and consistent clinical pulmonary function test to provide an indication of graft function" [71].

Table 1. Spirometry indices and definitions

Variables	Definition
FVC	Forced vital capacity; the total volume that can be forcefully expired From a maximum inspiratory effort (Liter)
FEV_1	Forced expiratory volume in 1 second; volume of air forcibly expired from a maximum inspiratory effort in the first second (Liter)
PEFR	Peak expiratory flow rate; the highest forced expiratory flow rate (Liter/second)
MEFR	Mid expiratory flow rate (Liter/second)
FEV_3	Forced expiratory volume in 3 seconds; volume of air forcibly expired from a maximum inspiratory effort in the third second (Liter)
FEF25	Flow rate at 25^{th} percentile of the FVC (Liter/second)
FEF50	Flow rate at 50^{th} percentile of the FVC (Liter/second)
FEF75	Flow rate at 75^{th} percentile of the FVC (Liter/second)

LTHMP

University of Minnesota Medical Center, Fairview (previously known as Fairview University Medical Center's or FUMC) is among the top frequented locations for lung transplant procedures in the United States [72]. The Lung Transplant Home Monitoring Program (LTHMP) [73], began in 1992, monitored a group of consented lung transplant recipients. More than 250 patients had participated in the research program.

The overall aim of LTHMP was to develop a low-cost tele- homecare system that would detect early signs of rejection or infection. One of the research interests is whether it is possible to react more quickly to disease changes for lung transplant patients by telemetric means. The researchers have attempted to detect onset of bronchopulmonary infections and rejections through data review.

In LTHMP, patients were trained to use a paperless electronic spirometer/diary at home to record and transmit daily vital signs, respiratory measurements, and symptom information. Typically, subjects would perform three forced vital capacity maneuvers at a consistent time in a day, preferably in the morning. The spirometer internally computes the values of FEV_1, FVC, MEFR and PEFR of each maneuver. The best performing spirometry (the largest combined FEV_1 and FVC values) was retained. This protocol ensures uniform adherence to American Thoracic Society standards of acceptability and reproducibility [71]. In addition, vital signs (pulse, weight, blood pressure, and body temperature) and respiratory symptoms (frequency of coughing/wheezing, amount and color of sputum, type and amount of exercise, shortness of breath at rest and during exercise, overall well-being, and stress level) were entered into the portable electronic spirometer by the subject. Subjects were instructed to complete a test session and transmit data back to the study data center [74]. Transmitted data were stored in a relational database. A sample home monitoring report is shown in Figure 5.

Study nurses at the data center reviewed data records each week and generated triage reports that identified patients who might have signs of clinical

problem. The transplant clinic staff would be notified to review subjects on the “watch list” and took appropriate course of action including contacting the patient and further deciding if medication change or an in-clinic evaluation was necessary. Patients were typically evaluated in clinic every two months during the first year, regardless of the data monitoring results. Patients may be seen more frequently if they complain about their symptoms or health. Patients were evaluated at a reduced frequency as they move farther from their transplant center or later date, depending on their conditions. Their clinic follow-up typically consisted of an annual visit, unless conditions suggested more frequent clinic contact.

The reliability and validity of the electronic spirometer/diary instrument used as a patient self-measurement device has been previously established [16]. However, it remains a challenge to maintain patients’ adherence to regular spirometer use and provide data [58]; often spirometry cannot be carried out in accordance with the conventional quality standards [71]; adherence deteriorates over time and becomes more of an issue for data integrity and completeness. To ensure good quality data, a great amount of effort was put into evaluating factors related to adherence and educating patients to maintain compliance to monitoring requirements [58,75].

Figure 5. A sample of the home monitoring report used in LTHMP.

Home Monitoring Report Mon Dec 22 00:00:00 2003 - Mon Jan 5 00:00:00 2004

Patient:5125:<->:xxxx

Date	Mn	FVC	FEV1	PFR	MFR	Mouthpiece	Wght	Temp	Plse	Sys	Dia	chg	SAmt	SCl	whz	sBr	sBx	xtm	xtp	wbg	str	Rx
BASELINE		33										2	1	2	0	0	0					
Mon Dec 22 10:27:00 2003	3	3.87	2.1	2.43	1.05	45754256	160	98.7	109	114	76	2	1	3	0	0	0	40	3	2	2	0
Thu Dec 25 10:16:00 2003	3	3.92	2.04	2.71	0.76	45754256	160	98	90	119	75	2	1	2	0	0	0	50	3	2	2	0
Fri Dec 26 08:54:00 2003	3	3.78	1.97	2.51	0.8	45754256	163	98	90	123	79	2	1	2	0	0	0	20	3	2	2	0
Sat Dec 27 10:03:00 2003	3	4.15	1.86	2.1	0.77	45754256	163	98	82	117	73	2	1	2	0	0	0	20	3	2	2	0
Sun Dec 28 10:06:00 2003	3	4.06	2.02	2.25	0.91	45754256	163	97.9	89	126	80	2	1	2	0	0	0	40	3	2	2	0
Mon Dec 29 08:35:00 2003	3	3.81	2.03	2.9	0.77	25754290	164	98.4	93	127	82	2	1	2	0	0	0	15	3	2	2	0
Tue Dec 30 09:02:00 2003	3	3.7	1.86	2.38	0.69	25754290	164	97.8	98	126	78	2	1	2	0	0	0	40	3	2	2	0
Wed Dec 31 09:32:00 2003	3	3.8	2.15	3.04	1.02	25754290	163	97.8	112	129	80	2	1	2	0	0	0	20	3	2	2	0
Thu Jan 1 11:20:00 2004	3	3.61	1.86	2.12	0.84	25754290	163	97.8	80	133	80	2	1	2	0	0	0	30	3	2	2	0
Fri Jan 2 10:13:00 2004	3	3.59	1.96	2.43	0.89	25754290	163	98.4	97	120	73	2	1	2	0	0	0	20	3	2	2	0
Sat Jan 3 09:01:00 2004	3	3.86	1.85	2.09	0.91	25754290	164	98.1	100	120	69	2	1	2	0	0	0	20	3	2	2	0
Sun Jan 4 09:19:00 2004	3	3.65	1.8	1.97	0.85	25754290	164	98.4	108	132	84	2	1	2	0	0	0	20	3	2	2	0
Last Week:	7	3.72	1.93	2.42	0.85	**2 Weeks:** 5		3.96	2	2.4	0.86											
		0.1	0.11	0.38	0.1			0.13	0.08	0.21	0.11											
3 Weeks:	7	3.99	2.09	2.67	0.89	**4 Weeks:** 6		3.78	2.08	3.42	0.85											
		0.15	0.07	0.17	0.07			0.17	0.08	0.37	0.12											
2 M.A.(0)	0	1M(14)	2.05	2M%	0	1M%	**-0.06**	4W%	**-0.07**	3W%	**-0.07**	2W%	-0.03									

Home Monitoring Report

Page 1

Pulmonary related symptoms have been the common reasons that trigger a clinic visit and play an important defense role for patient safety. Symptom reporting is also the least invasive type of surveillance to capture events [76].

Symptom data are subjective in nature. The following Table 2 illustrates the numerical coding of the respiratory symptom variables in LTHMP that include cough, wheeze, sputum (amount, color) and shortness of breath.

Table 2 . Symptom data scales: cough, wheeze, sputum (amount, color) and shortness of breath (dyspnea)

Symptom\ score	0	1	2	3	4
Shortness of breath* (SBr)	None	Mild	Moderate	Severe,	
Wheeze (Whz)	0-2 /day	Few hrs	Several/hr	Many/hr	
Cough (SBx)	0-2 /day	Few hrs	Several/hr	Many/hr	
Sputum (SAmt)	None	Small	Moderate	Large	
Sputum color (SCl)	None	Watery	White	Yellowish	Dark yellow/blood tinged

Note: Abbreviations as in the sample home monitoring report (Figure 5)
*: SBr at rest after exercise.

Event Detection in LTHMP and Related Fields

Qualitative Assessment

Some related research resorts to visual display and simple heuristics to process home monitoring variables [77]. Information display formats including graphical, table, hybrid (combination of graphics and table), and control chart presentations were compared for effectiveness for decision-making. Such methods themselves do not provide quantitative assessment of the home monitored data.

Rule Based Decision Support

The Triage Algorithm

In LTHMP, spirometry and symptom data were reviewed weekly by research nurses following expert derived guidelines known as the triage rules. Computer generated data summaries were used to assess data trend and compare to

previous data points. Human data review typically suffers from subjectivity, stress, fatigue, and expensive labor cost, conclusions can also differ depending on the reviewer. In LTHMP, increasing number of transplanted patients and expanding data volumes had increased burden to the data monitoring staff, yet the resource is relatively unchanging. Hence, one top priority of the earlier research was to reduce the data burden imposed on the staff. An automated nurse triage algorithm had been developed.

The basic idea of the computerized automated triage algorithm [17] was to compare the current week's average with the maxima of the preceding weekly or monthly averages. Three levels of decline thresholds were selected based on the starting baseline FEV_1. Similarly, the rules include heuristics based on the increase of existing or emergence of new respiratory symptoms. If the severity condition is met, a “watch alert” will be issued.

The automated triage algorithm has been validated against the manual triage review and therefore, is considered an “optimal” replica of the manual triage. The design goal was to identify patients who should be watched for potentially having developed respiratory-related problems, rather than those who were actually experiencing an event. Because the algorithm design was not based on actual events, it was not valid for use in event detection. The main reasons for not aiming at actual event is that, to ascertain accurate event annotations would require dedicated professional to carefully review the paper based medical records which later transitioned into an electronic record system.

One drawback of the rule-based logics is that the binary decision can not be easily quantified as the likelihood of event.

Event Detection Using Bayesian Statistics

Bayesian statistics have gained more popularity in the last decades due to that computation intensive solutions are widely available with dedicated software packages [78,79,80].

In LTHMP, event detection based on Bayesian statistics has been explored [20]. Detection was performed in an off-line setting where “epochs” - a set of consecutive data points of two-week length were selected from the dataset including half of them that end with events and half that end without events, form binary trials. This translates to approximately 10% of the data (150 two week epochs over 60 patient-years) utilized for training and testing. Bayesian statistical models, including a step function for “mean change” and a linear regression for “slope change” models were fit to each epoch and the probability of event was estimated and served the basis for event detection.

Because these Bayesian models were trained and tested on a fixed number of non-overlapping epochs, the scheme is a fixed sample, trial-based detection. Each epoch is assumed to have a fixed duration with known event outcome, whereas in clinical setting, event detection must perform without prior knowledge of the epoch duration or outcome. Furthermore, because the models are based on the entire epochs, timing of detections could not be assessed.

Detection of Change

These time domain methods primarily concerns trend analysis in the original data format. In biomedical applications, a trend is seen as a general direction of the mean level in a set of data [81]. Trend analysis involves examining time series data and identifying significant increases or decreases in the magnitude of a reference variable [82]. Event detection is concerned about the change in trend.

Statistical Detection of Change

One major theoretical advance in the 1950s was to combine detection theory with statistical decision theory [83]. A well known problem is the change point detection [84].

In general, assume a sensory response that reflects the underlying stimuli. If the signal is present, data observation is a combination of signal and noise. In its

basic form, assuming a noise distribution and a signal-plus-noise, the detection is performed by a statistical hypothesis testing:

H_0(Null hypothesis): $\mu = \mu_0 = 0$ (noise only)

H_1(Alternative hypothesis): $\mu = \mu_s$ (signal and noise)

When the cumulative probability distribution function (PDF) of μ_s crosses a specified threshold, the null hypothesis is rejected in favor of the alternative, i.e., the signal is present. The previously mentioned Bayesian statistics can be considered a special case in change detection.

Sequential Test for Change

A natural setting for event detection in real time is the sequential probability ratio test (SPRT) [85], a sequential hypothesis test developed by Wald [86]. As in the classical hypothesis testing, SPRT starts with a pair of hypotheses, H_0 and H_1 for the null hypothesis and alternative hypothesis respectively. The next step is to calculate the cumulative sum of the log-likelihood ratio, $log\Lambda_i$, as new data arrive:

$$S_i = S_{i-1} + log\Lambda_i$$

The stopping rule is a simple thresholding scheme:

- $a < S_i < b$: continue monitoring (*critical inequality*)
- $S_i \geq b$: Accept H_1
- $S_i < a$: Accept H_0

The main challenge in statistical detections is to correctly define the signal to be detected and its underlying probability distribution, and modeling data traces at the individual subject level.

Signal Processing Approaches

Signal processing methodologies depend on the applications. Many complex signal analysis problems are solved using frequency analysis such as Fourier analysis.

Certain event of interest is detected based on known signal characteristics including waveform shapes and frequency signature, such as seizures showing as large discrete spikes, absent from a normal EEG in the electroencephalogram (EEG) analysis [87,88].

Many signals show no organized patterns. For instance, in gait control studies which measure the walking patterns to detect orthopedic impairment [89] and in heart rate variability analysis of cardiac health [90,91], the impaired or disease subjects are manifest with different level of regularity than the normal subjects.

For time-dependent events, static features that are not localized in time would not be useful as they do not provide insight to data dynamics related to events. The context of event is also confined in regard to the time frame relevant to the event and the amount of change that can happen within that time frame. In internet traffic monitoring, detection of malicious network intrusions can benefit by re-projecting different locally observed data streams into alternate coordinates by using Principal Component Analysis, the components associated with the normal traffic and components associated abnormal behaviors could be separated [92, 93].

Thus, features to be extract from the data should reflect the time frames in which the event of interest develops. The analysis of data correlation across multiple time scales can identify data patterns related to change, either temporal (in the time domain), or in the frequency domain, or in both.

Chapter III

MATERIALS AND DATA SETS

This chapter first describes the datasets that were used as the basis to learn, explore and evaluate various analytical approaches, followed by the dataset used to validate the proposed ideas. Characteristics of the data are illustrated with cases.

The Learning Set

The Lung Transplant Home Monitoring Program (LTHMP) at the University of Minnesota provided the home monitoring data. In LTHMP, patients were trained to use an electronic spirometer/diary device at home to record and transmit daily respiratory measurements, vital signs (blood pressure, heart rate, temperature, and weight), and symptoms [94]. Home monitoring records from 28 consecutive subjects with adequate length of follow-up (>60 days of spirometry monitoring) [20] were used. The dataset contains patient identifier, gender, date of birth, primary indication for transplant, date of transplant, the type of transplant, dates of the monitoring, home spirometry involving variables $FEV_{1,}$ FVC, MEFR and PEFR, daily temperature, and daily symptoms (cough, dyspnea, wheeze and sputum in ordinal scales). Data were recorded daily except for occasional gaps where subjects did not record or transmit data. The mean age at the time of transplant was 50.8 years (±SD 12.5, range 22.5 - 66.5 years) and 10 (36%) were female. There were 17 single lung, 10 bilateral single lung, and one heart-lung transplants. Home monitoring started from 26 to 367 days post transplant. There were a total of 10,098 samples of daily records representing a total of 13,609 calendar days or 37.3 subject-years of data where the difference in total days reflects the gap of days when home testing was not performed/ transmitted. The mean monitoring time was 1.3 years/subject (469 days/subject) (±SD 228 days or 0.8 year, range

60 days - 2.8 years). All subjects provided written informed consent to participate in LTHMP following the University of Minnesota Institutional Review Board guidelines.

Table 3. Demographic characteristics of subjects

Characteristic	No. of subjects (N=28)
Gender	
Female	10
Type of transplant	
Single	17
Double	10
Heart-lung	1
Pretransplant diagnosis	
Chronic obstructive pulmonary disease (COPD)	14
Idiopathic pulmonary fibrosis (IPF)	5
Cystic fibrosis	3
α_1-Antirypsin deficiency	2
Bronchiectasis	2
Other*	4

* Pulmonary hypertension, bronchiectasis, Eisenmenger syndrome, lymphangioleiomyomatosis

Events

Events of interest were acute bronchopulmonary related clinical complications such as bronchitis, pneumonia, pneumonitis, acute rejection, and newly diagnosed or treated episode of an acute, primary bronchopulmonary infection or rejection which is accompanied by the start of a new treatment or the change of a chronic medication [95]. The determination of events was based on a careful review of the patients' medical records for documentation of adverse effects and treatments, and biopsy records, or both. Events were determined by a qualified health science professional who retrospectively reviewed the medical records of each subject. The event information included patient identification number and documented date of the event(s). Hospitalization and telephone prescription records were not part of the event review. Events do not explicitly distinguish between infections and rejections. This research includes 10 additional events from transbronchial lung biopsies with graded acute rejection and infections which were maintained in a separate database [96]. This resulted in a total of 101 acute bronchopulmonary related events in 26 subjects. Events/subject ranged from 1-8 events, and the median was of 2 events /subject. The remaining 2 subjects did not have events during the monitoring period.

Data Illustration and Characteristics

The following illustrates a set of representative home monitored data series from the learning set. Spirometry and symptom series are displayed within two time frames, first by a shorter set including the first 120 days (Figure 6 and Figure 7), then the entire set of the spirometry up to 800 days (Figure 8). In this case, two events were reported on day 97 and day 267.

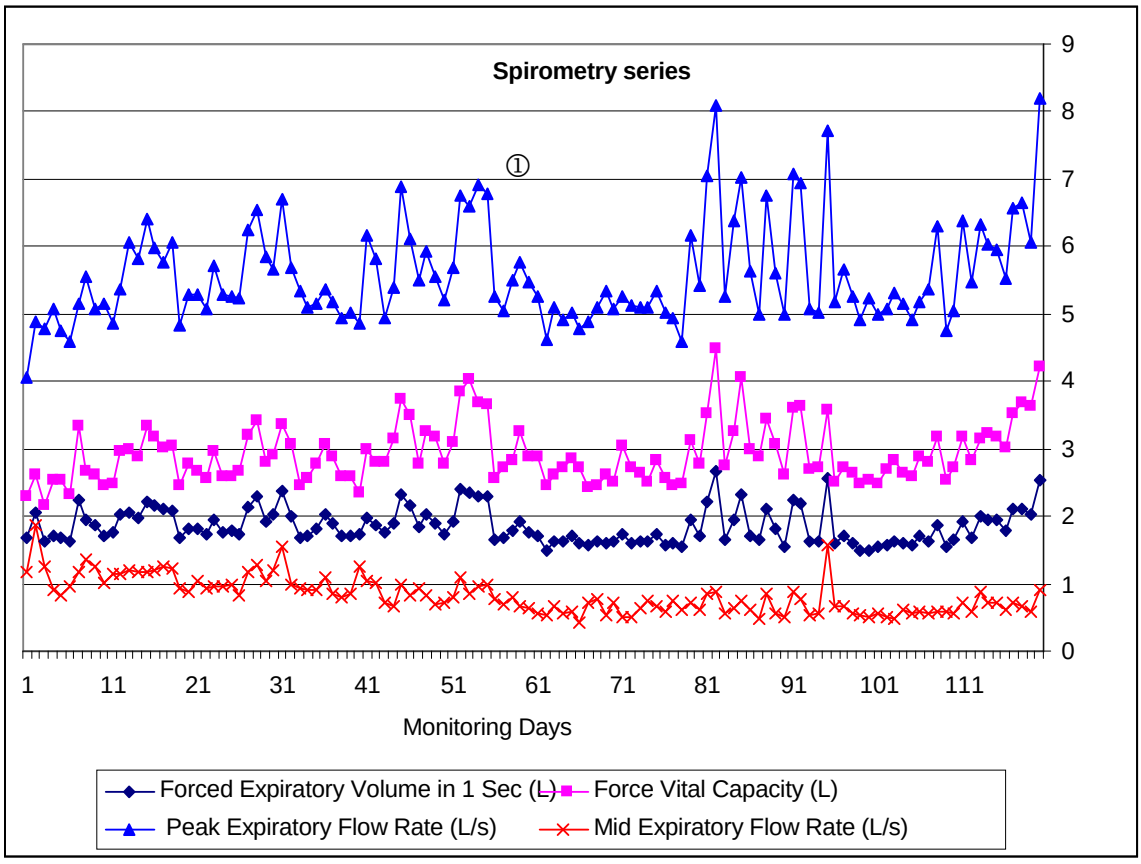

Figure 6. Illustration of the spirometry series

(Subject X5198)

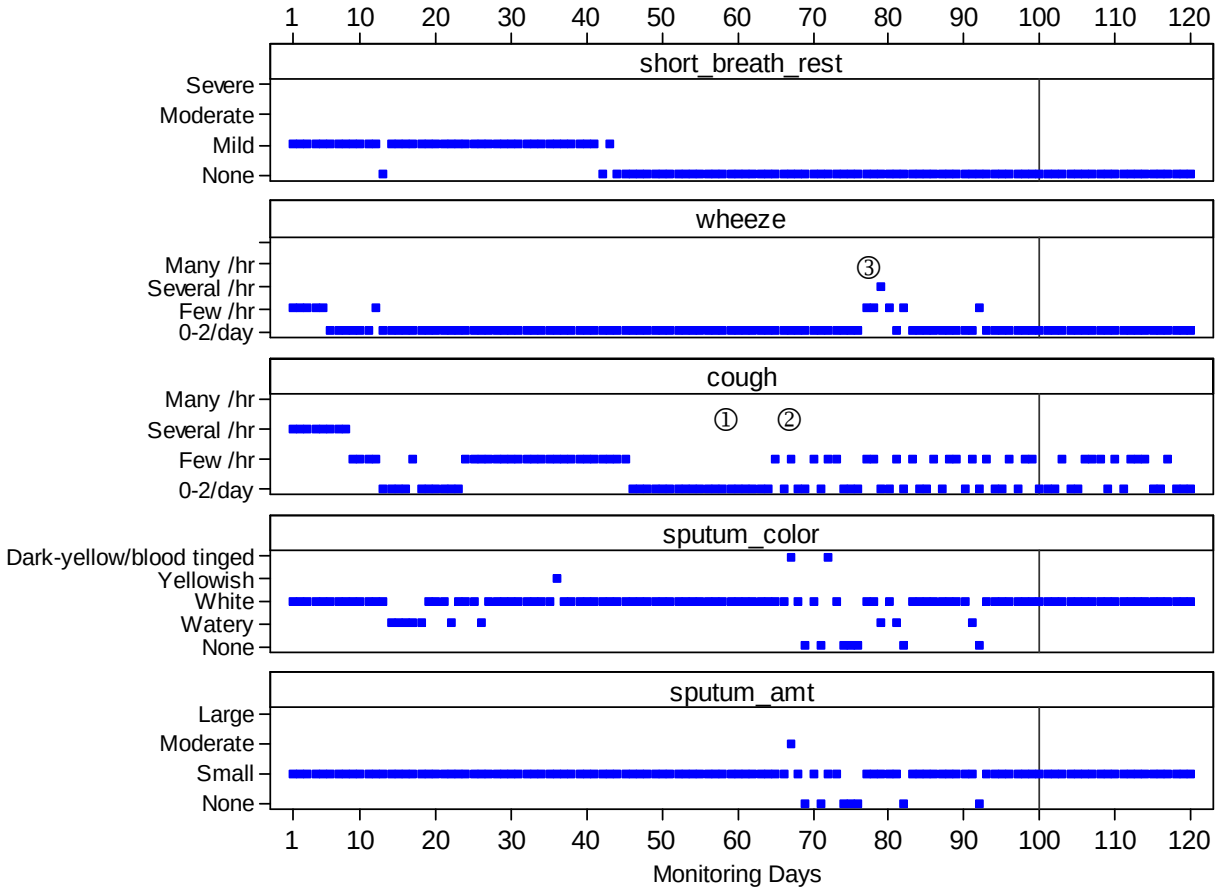

Figure 7. Illustration of the symptom reports

(Subject X5198)

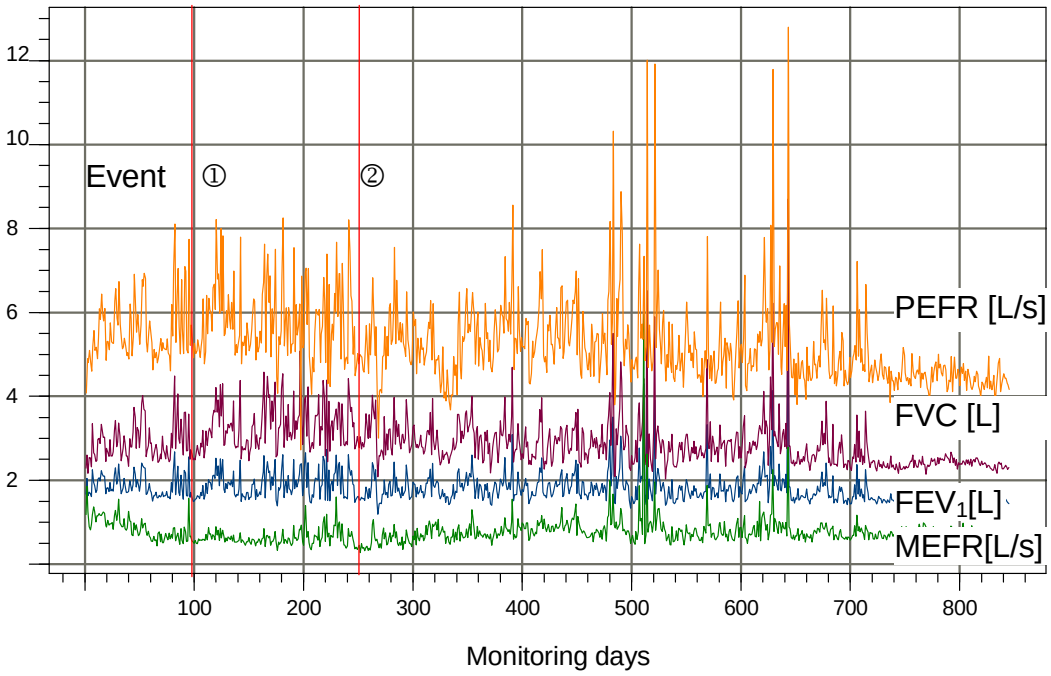

Figure 8. Illustration of the long term spirometry time series

Horizontal axis is the days of monitoring, start from the first day of home monitoring. Vertical axis is according to the unit of each variable.

(Subject X5198)

As illustrated in Figure 6 - Figure 8, a few important data characteristics are observed:

Home spirometry data series presents a wide range of inter-day variability. Data are often highly fluctuating, do not look smooth and with sharp transitions (discontinuity), and appear noisy. Some subject's data fluctuate significantly throughout the course of the monitoring. Thus, both the mean and variance could change from time to time. Such changing variability and complex patterns are best described as nonlinear which can not be represented by a linear function, and nonstationary whose statistical property such as means and variance change over time. There appears to be no periodicity. There is also a self-similarity between signals in a fine time frame and a coarse time frame. This "fractal" time-scaling behavior is another indication of nonlinearity.

The data variability may be contributed by numerous factors: the spirometry test itself, per definition, is effort dependent; there are human errors while performing forced expiratory maneuvers such as air leaks; there are potential influence of environmental changes such as the barometric atmosphere pressure; as well as factors that can not be enumerated for each maneuver circumstance.

The timing of data variability in spirometry and symptom variables does not always align: for example, spirometry showed depressed values starting around day 60 (Figure 6①,); while coughing and sputum were not manifest until around day 70 (Figure 7②), and wheeze started around day 80 (Figure 8 ③). In this example, spirometry leads the symptoms in showing earlier signs of a problem.

Based on the learning set, there are different degrees of variability from one event to another event, and from one subject to another subject. Data variability associated with events in one subject can be very similar to that in another subject without an event. There is a wide range of intra- subject and inter- subject variability. As a result, events are not easily visible to the naked-eye.

The autocorrelation plots [97] indicate that: all spirometry variables exhibit long range dependence; the data series has deterministic component and is not random. In particular, the autocorrelation in FEV_1 is lower than other variables at the same lag parameter, suggesting FEV_1 requires more frequent sampling to capture the dynamical information.

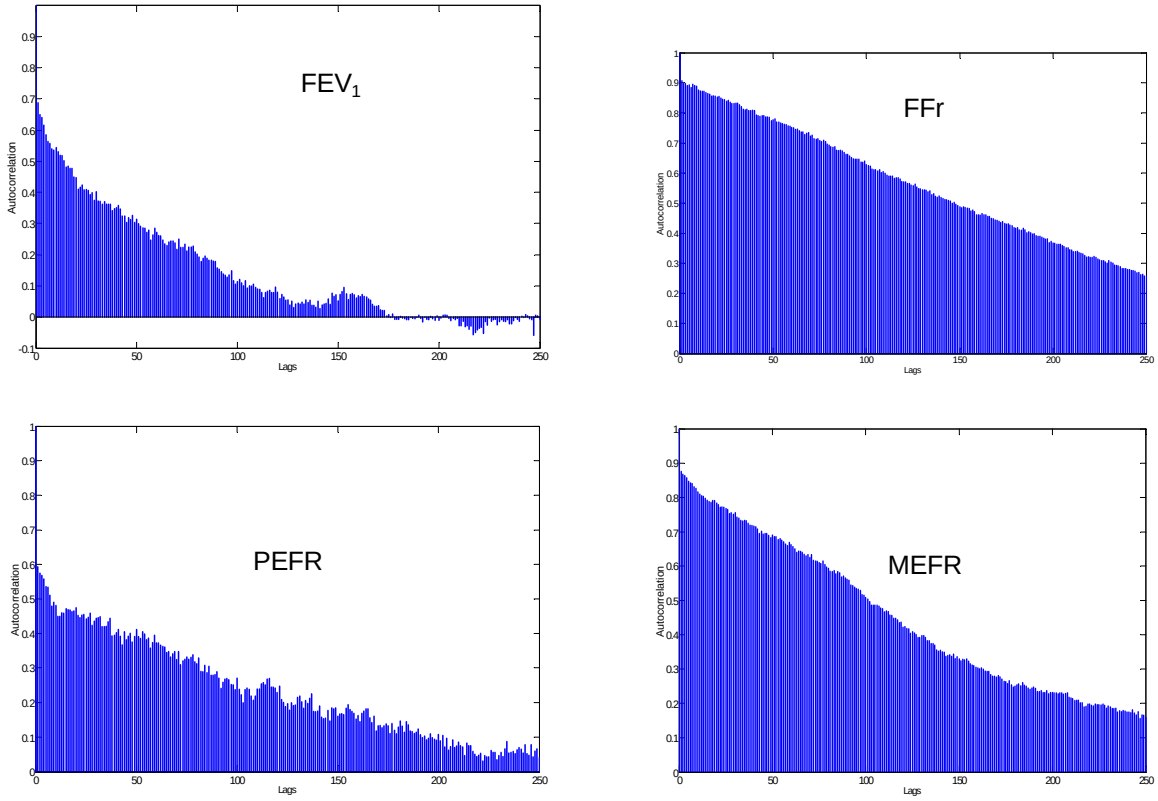

Figure 9. Autocorrelation plots

(Subject X5183)

In summary, home monitoring data series are complex with multiple characteristics: data are not smooth, nonlinear, nonstationary and autocorrelated long range dependence; signs of events based on spirometry and symptom series are not necessarily manifest at the same time; and it would be challenging to accurately divide the series into uniform-sized “epochs” as binary trials.

Data Pre-processing

LTHMP has its own data screening and correction mechanism, and the number of totally erroneous data points is low. No deliberate attempt was made to correct data points. However, data series were scrutinized for transcription errors and data omissions.

Data Omissions

These data in clinical setting, collected and reported by the patients themselves contain unequally spaced data points. A close scrutiny of the learning set reveals a wide range of monitoring completeness: from one subject reporting 301 out of 662 days (45%) to another subject reporting 350 out of 352 days (99%). The following table summarizes the data omissions in the learning set.

Table 4. Summary of data completeness in the learning set.

	Mean	±Std Dev	Min	Max	Median
Total Sample days	361	201	60	1023	306
Total calendar days	486	244	107	1421	416
Ratio of reported days/ total calendar days (sample/total)	74%	16%	45%	99%	75%

Spirometry Data

For spirometry data, rare physiologically implausible outliers were first spotted and removed. The data indexing was maintained according to the calendar date.

Symptom Data

In symptom data, there were occasional omissions in one or more of the symptom frequency-severity categories for a particular day (typically less than 5% of available data). These missing values were replaced with the previous record (“last observation carried forward”) which typically contained no new information. However, this would permit mathematical operation such as summation on the available data without resetting the entire symptom categories to be missing for that day.

Symptom reporting consists of frequency and severity of cough, dyspnea, wheeze, sputum amount and sputum color. Unlike objectively measured spirometry data, symptom reports are subjectively self-reported in ordinal scales (Table 2). Thus, the domain expert defined heuristics are highly valued as they provide the best insight how to interpret the subjective data. The necessary step is to convert these symptom scores into a composite score.

By emulating the automated triage rules developed previously [17], the original symptom variables of dyspnea, wheeze, sputum, and cough were converted into one composite via weighing and summation. The weight assignment emulates the expert defined rules [17]. For example, the rule suggests weighing newly reported symptoms of dyspnea or wheeze equally to two degree increase in sputum or cough (e.g., from small to large amount). A moving symptom baseline is calculated as the average of each symptom level over an 8-week period starting 9 weeks before the current weekly report period and ending 1 week before the current report period.

Table 5. Expert defined important thresholds (also see Table 2)

Symptom	Level threshold	Level change threshold
Cough	--	2 degree increase
Sputum amount	3 (large)	2 degree increase
Sputum color	3 (yellowish)	2 degree increase
Wheezing	1 (presence)	--
Dyspnea at rest	1 (presence)	--

Notes: The level threshold is the minimum value of each symptom that will fire the alarm. The level change threshold is the minimum value of the difference between the current symptom value and the patient's baseline symptom value that will fire the alarm.

The expert rule weighs the presence of wheezing and dyspnea equal to significant sputum or the 2 degree changes in cough and sputum over the baseline. Therefore, the composite symptom score is

$$Symptom = \sum a_i x_i + \sum b_j y_j \ , \qquad (1)$$

where

$a_i = 1$ for x_i= wheeze, dyspnea

$a_i = \frac{1}{3}$ for x_i=sputum amount, sputum color,

$b_j = \frac{1}{2}$ for y_j = Δ (change) in cough, sputum amount, sputum color, over the baseline, respectively.

Since symptom scores are bounded by their permissible ranges, the summation of the scores are also bounded without extreme values. The resulting symptom composite series resemble a near continuous measurement series, illustrated below.

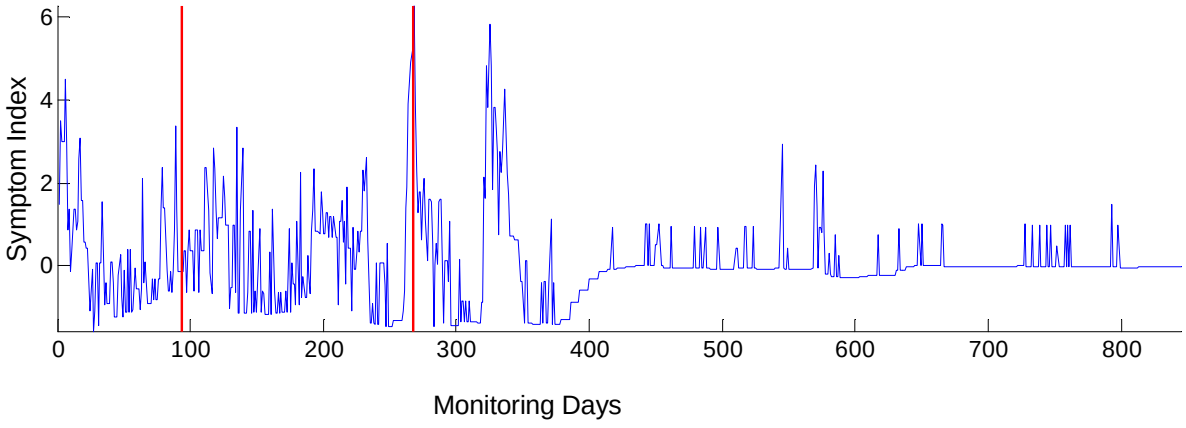

Figure 10. Illustration of the converted symptom index

(Subject X5198)

The Validation Set

For validation purposes, home monitoring records from 30 different subjects between April 2007 and December 2007 that were not part of the learning set were disclosed to this research after the learning stage. This constitutes a new validation set. All subjects provided written informed consent to participate in the home monitoring program following the University of Minnesota Institutional Review Board guidelines.

The format of the data is similar to the learning dataset. No additional patient demographic information was disclosed. This includes a total of 2,426 samples of records representing a total of 4,625 calendar days or 12.6 subject-years of data where the difference reflects when home testing was not performed. The average duration of monitoring was 148 days per subject (±SD 66 days, ranged from 24 - 269 days).

During 2007, an event diary was actively maintained by a dedicated study nurse, who followed up with physicians outside FUMC to ascertain complete record of events, which provides the basis of event status. A total of 40 acute bronchopulmonary events were identified in 14 subjects (range 1-7 events/subject, median 2 events), and the remaining 16 subjects did not have events during the monitoring period.

Chapter IV

WAVELET METHODS FOR EVENT DETECTION

Wavelet Methodology

Home monitored data series from LTHMP exhibit dynamical nonlinear non-stationary characteristics with time-scaling behavior. Wavelet methods, developed by Mallat [98], Meyer [99], Daubechies [100,101], and Vaidyanathan [102] are well-suited for such type of data.

Wavelets, meaning “small waves”, are a set of functions which are obtained by translating and contracting or dilating a prototype wavelet, which is an orthogonal, finite energy signal with compact support. This way, a wavelet transformation decomposes a signal, simultaneously showing its time-scale (resolution) representation. Intuitively, the wavelet methods are described as “seeing both the trees and the forest”, implying the analysis of the same signal at different resolutions [103,104,105].

Tewfik and Kim [106] were among the first to discuss the decorrelation property of wavelet transforms for fractional Brownian motion. This decorrelation property is critical to perform scale dependent thresholding, known as wavelet shrinkage, proposed by Donoho and Johnstone [107] to reduce noise. Wavelet shrinkage is directly related to nonlinear regression and can lead to an optimal data recovery from noises [108]. Wavelet based multiresolution decomposition has proven to be a very effective tool to analyze nonstationary signals [109,110,111], particularly when noises are correlated. Although best known for image processing and data compression applications, wavelet has evolved over the years to be very useful in signal processing, trend analysis and data mining [112] and has thus been widely used in the fields of physiological and biomedical signal analysis to characterize signals [113].

The discrete wavelet transform (DWT) analyzes the signal at discrete resolutions through decomposition of the signal into successive frequency bands (hence, multiresolution analysis). It is more convenient to restrict τ to dyadic scales $\tau_j=2^j$, $j=0, 1, \ldots, J$. The DWT utilizes a pair of functions, a wavelet function $\psi(t)$ and a scaling function $\phi(t)$ where these functions are a weighed sum of the scaled (dilated or contracted) and shifted version of the scaling function itself:

$$\phi(t) = \sum_n h[n]\phi(2t-n)$$

$$\psi(t) = \sum_n g[n]\phi(2t-n)$$

where a dyadic discrete wavelet scaled by 2^j is defined as [114]

$$\phi_{u,j}(t) = \frac{1}{\sqrt{2^j}}\phi(\frac{t-u}{2^j}),\ j \in Z\ ,$$

and

$$\psi_{u,j}(t) = \frac{1}{\sqrt{2^j}}\psi(\frac{t-u}{2^j}),\ j \in Z\ .$$

and restrict time points to integers $u=1, 2, \ldots, N$.

Applying DWT to discrete-time signal series $x(t)$, $t=1,\ldots,n=2^J$, creates a multiresolution decomposition of $x(t)$,

$$DWT(x(t))=(a_J|d_J|d_{J-1}|\ldots|d_2|d_1)$$

of the same length n, on j octaves labeled by $j=1, \ldots,J$. a_J contains the low frequency component of the signal, referred as wavelet approximation coefficients, and $d_J|d_{J-1}|..d_2|d_1$ contains wavelet detail coefficients at decreasing dyadic scales.

Further,

$$x(t) = A_j(t) + \sum_{j=1}^{J} D_j(t),$$

$$x(t) = \sum_{k \in Z} a_{J,k} \phi_J [t - 2^J k] + \sum_{j=1}^{J} \sum_{k \in Z} d_{j,k} \psi_j [t - 2^j k].$$

where $a_{u,j}$ and $d_{j,k}$ are the DWT coefficients and

$$a_{J,k} = \int x(t) \phi_{u,j}(t) dt, d_{j,k} = \int x(t) \psi_{u,j}(t) dt \text{ for } j=1,\ldots,J-1; k=0,\ldots,2^j - 1.$$

In summary, the DWT signal estimation can be summarized by a three-step procedure:

$$x \xrightarrow{DWT} \{\hat{a}_J, \hat{d}_{j,k}\} \xrightarrow{\text{coeff. processing}} \{\hat{a}_J, \hat{d}'_{j,k}\} \xrightarrow{IDWT} \hat{f}_\lambda$$

which consists of decomposition process, coefficient processing, and inverse DWT (IDWT) reconstruction.

Event Detection Steps

The event detection algorithm development involved the following steps: wavelet-based signal analysis; feature discovery/ extraction; performance evaluation; and validation [115].

The critical step of feature extraction contains a reduced representation instead of the full size data input. The middle layer features (between original data and decisions) provide insight as to when event(s) may have occurred and serve the basis for event detection.

Wavelet Analysis of the LTHMP Data

Choice of Wavelets

In this research, Daubechies wavelets family was used. This choice of wavelet is based on the principal of “parsimony”. That is, Daubechies wavelets are orthonormal and compactly supported with few elements in the basis functions and near symmetrical; the shape of the basis function resembles the time series signal display in LTHMP which is not curvilinear smooth and consists of sharp jumps and discontinuities. In fact, in estimation of rapidly changing non-stationary signals, basis functions are best selected from what resembles the signal itself [116].

The Daubechies wavelets are denoted as db*N*, by the highest number of vanishing moments N, db3 (top, Figure 11) was used to perform a 3-level wavelet decomposition (higher resolution); db4 (bottom, Figure 11) was used to perform a 6-level wavelet decomposition (lower resolution).

As illustrated, the Daubechies wavelets visually resemble a local data display of the spirometry series. This shape similarity is one of the reasons Daubechies wavelets have been widely used to fit a nonlinear nonstationary signal series with discontinuities.

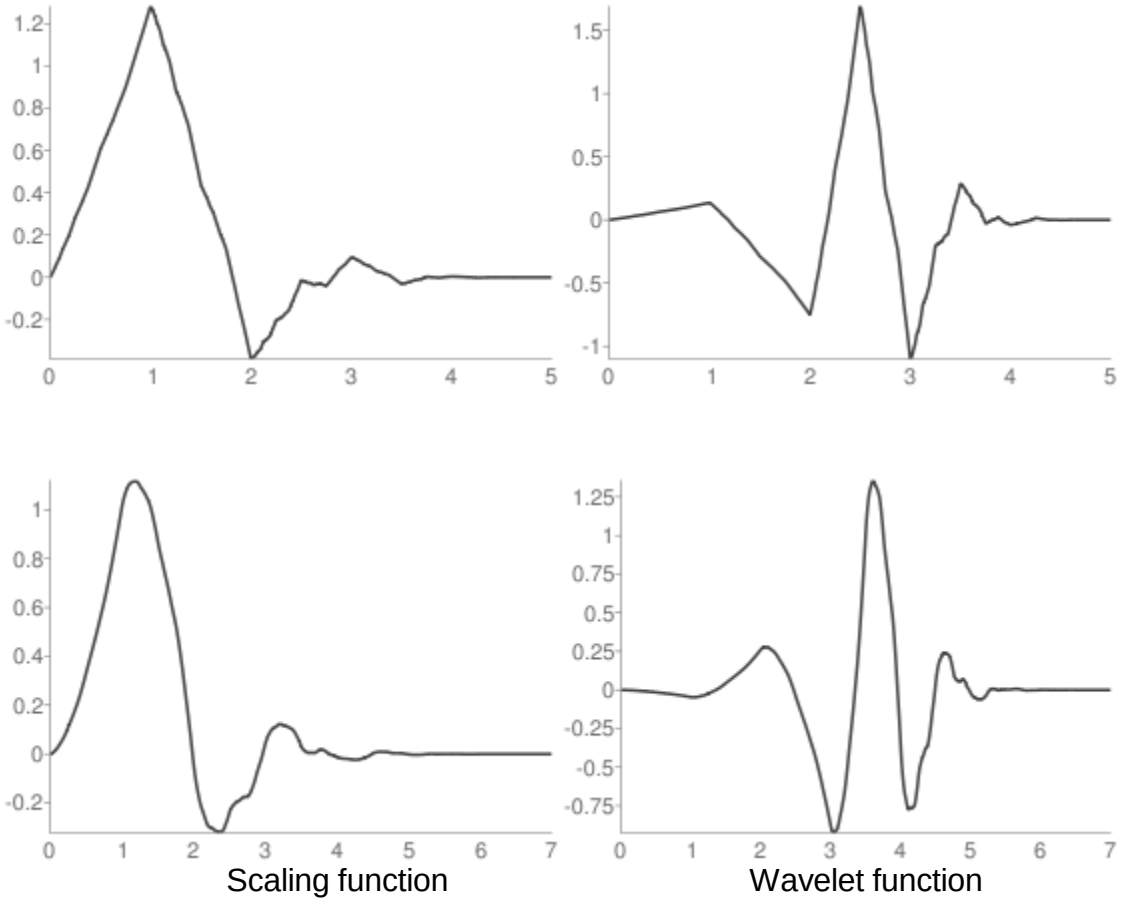

Figure 11. Daubechies wavelets db3 (top) and db4 (bottom)

(Source: [117])

Wavelet analysis was applied to LTHMP data to re-estimate signal amid noise.

Let y_i be the data series from one subject in LTHMP, at indexed time t_i, to recover the true underlying signal $f_i, i = 1,2,...,n$ from

$$y_i = f_i + \varepsilon_i \tag{2}$$

with noise ε_i. Under the matrix representations, let

$w=\boldsymbol{H}y,\ \theta=\boldsymbol{H}f,\ v=\boldsymbol{H}\varepsilon,$

where $\boldsymbol{H}$ is the dyadic orthonormal wavelet transformation matrix, let θ_i^j be the wavelet detail coefficient of the true signal f_i at time i, scale j and w_i^j and v_i^j defined similarly for the observed data and the noise, respectively. Due to the orthogonality of the DWT,

$w_i^j = \theta_i^j + v_i^j$.

The wavelet approximation coefficients were kept intact as they represent low-frequency deterministic component representing the temporal trend in the signal.

Let w_i =(w_1,w_2,…,w_n)' be the wavelet coefficients from $w= \boldsymbol{H}y$, and let the noise term $\varepsilon_i = \sigma \cdot z_i$, where z_i is independent Gaussian random noise with mean zero and variance one, one computed an estimate of σ, the standard deviation of the noise from the wavelet coefficients w_i .

A shrinkage rule $\delta_T(.)$ ("WaveShrink") [118] was then applied to the coefficients of w_i. Each wavelet detail coefficients w_i was rejected or retained into $\hat{w}_i$ per threshold T

$$\delta_T(w_i) = \begin{cases} 0, & if\ |w_i| \le T \\ sign[w_i](|w_i| - T), & if\ |w_i| > T \end{cases} \tag{3a}$$

The Universal Threshold T [119] is defined as

$$T = \sigma\sqrt{2\log(n)}\,, \tag{3b}$$

where n denotes the signal length, and σ denotes the estimate of the standard deviation of noise, given by

$$\sigma = \frac{MAD}{0.6745}\,, \tag{4}$$

where the MAD is the median absolute deviation of the wavelet coefficients estimated in the first scale and the factor 0.6745 is used for calibration with the standard Gaussian distribution. The universal threshold method involves less computing cost, relatively simple, and would produce a “noise-free” estimate as for a wide range of n, the expected number of coefficients exceeding the threshold is relatively low and T is optimal with random Gaussian noise [120].

Let θ be the noise free true wavelet coefficients, under the orthogonal wavelet transform, one has

$$w = N(\theta, \sigma^2 HH^T) = N(\theta, \sigma^2 I_n)\,,$$

Because of orthogonality of $\boldsymbol{H}$, the inverse DWT (IDWT) is simply given by $f = \boldsymbol{H}^{\boldsymbol{T}}\theta$. Therefore, by thresholding or shrinking the detail coefficients and inverting the DWT, we eliminated v and obtained $\hat{f}$, an estimate of the underlying signal f.

the “WaveShrink” estimate can be written as

$$\hat{f}_i = \sum_k h_{ki}\hat{w}_k, i = 1,2,...,n; k = 1,2,...,n \tag{5}$$

and

$$E(\hat{f}_i) = \sum_k h_{ki} E(\hat{w}_k)\,,$$

$$\mathrm{var}(\hat{f}_i) = \sum_k h^2{}_{ki}\ \mathrm{var}(\hat{w}_k)$$

The indexing of sequential data points is maintained by t_i, $i=1,2,\ldots, n$.. It is necessary to preserve the correct data spacing so that the time-scale relationship is protected. The basic approach is to map unequally spaced data to a regular grid [121,122], and the unobserved consecutive data points were linearly interpolated between two known data points. Then wavelet shrinkage and reconstruction per the inverse wavelet transformation were performed as previously described. Data indices corresponding to the interpolation were tracked and event detection would not occur at these indices.

Illustration of Wavelet Method

The wavelet decomposition overcomes the issues of autocorrelation which results in nearly uncorrelated components in the wavelet space.

Two figures (Figure 12 and Figure 13) illustrate such effect. First, a four-level DWT was performed on the FEV_1 data. Figure 12 shows the original data series, followed by the DWT decomposition. FEV_1 was decomposed into approximation A4 and detail coefficients D1, D2, D3, and D4. The lower numbered scales give the details corresponding to the high frequency components of the signal and the higher scales correspond to the low frequency components. DWT identifies those larger coefficients that are most significant for representing the data, which are to be separated from noise related small coefficients.

Figure 13 shows the approximation coefficients (A4) has slowly decaying autocorrelation, which represents the deterministic component in the data. The autocorrelation in the wavelet detail coefficients in each scale quickly decays and becomes nearly uncorrelated (shown in D1-D4). This decorrelation property is critical as this is the direct result of the decomposition process.

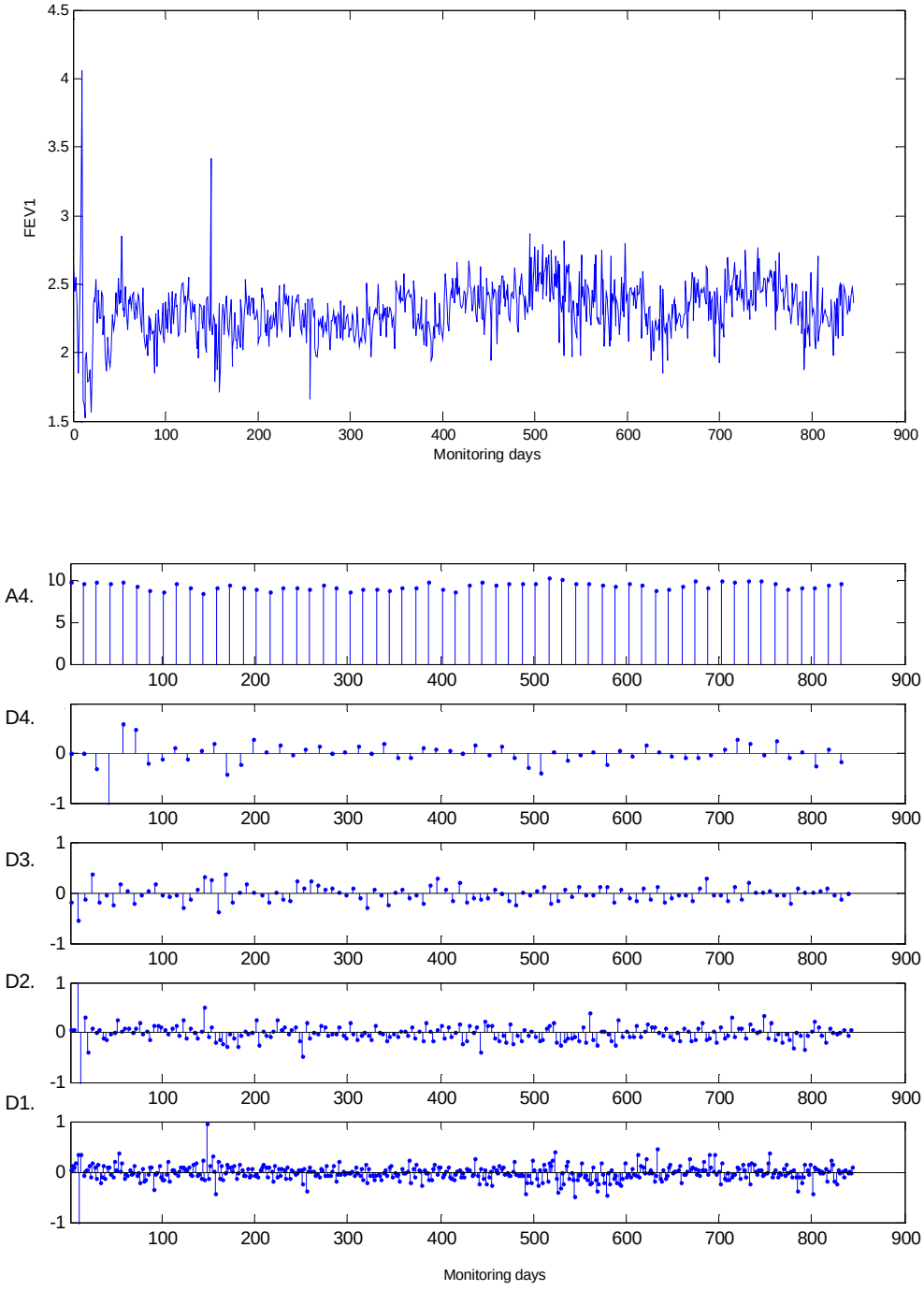

Figure 12. Data series FEV_1 and its four-level decomposition

Each vertical bar indicates the position and amplitude of the corresponding wavelet coefficients. Scales are 2^{J-1}, J=1,2,3,4. (Subject X5183).

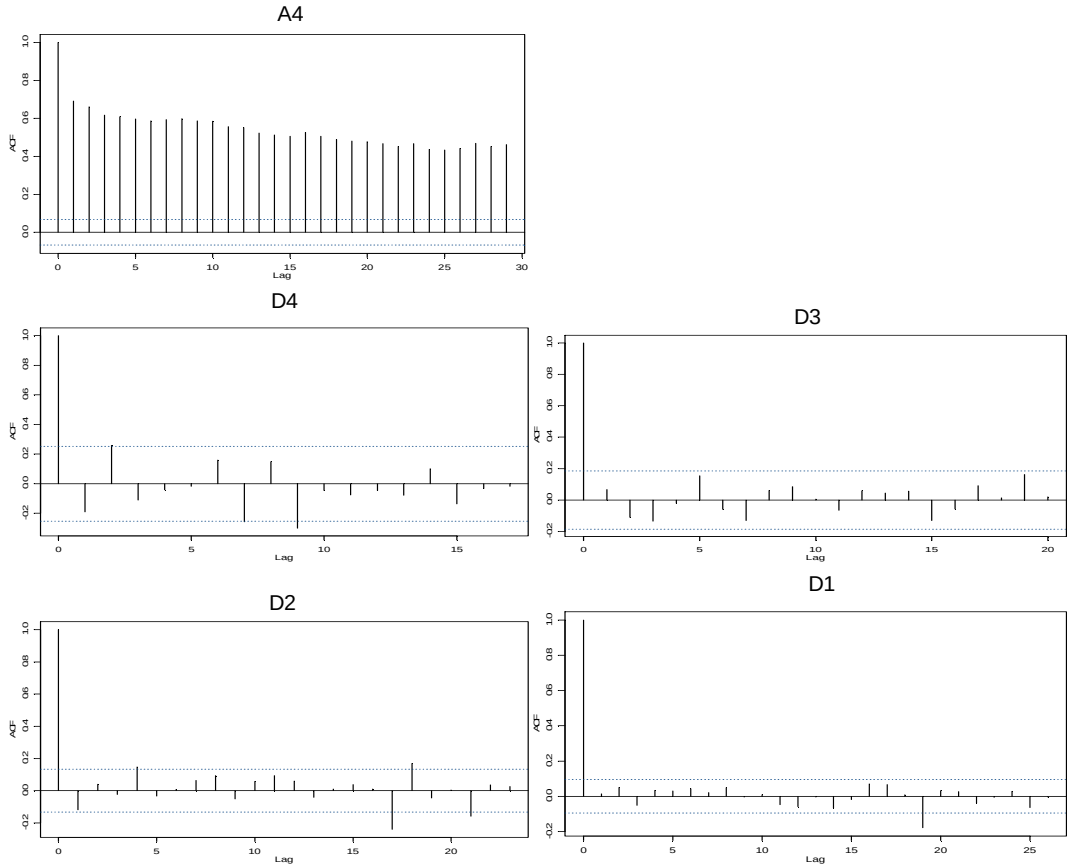

Figure 13. Autocorrelation function (ACF) - lag plots of wavelet coefficients.

ACF-lag plots of A4, D1-D4 at scale 2^{J-1}, $J=1,2,3,4$. (Subject X5183).

Wavelet Improves Signal Noise Ratio

The wavelet method provides excellent noise removal choices by retaining only significant wavelet coefficients. Inverse wavelet transformation reconstructs from the reduced wavelet space as the de-noised signal estimate. This is illustrated next.

Data estimation by wavelets and by a moving average of a sample FEV_1 series is shown in Figure 14 [Panel A]. The signal noise ratio (SNR) was computed and compared ([Panel B]) with a sliding window size $n=20$ and an updating block size $n_x=2$. It is shown that SNR by wavelets(DWT) is consistently higher than that of a moving average approach (MA). The wavelet estimate closely traces the data points, with no phase shift, and it is not over influenced by noisy outlying observations. The moving average estimate provides a reasonable fit but it noticeably deviates from the center of data points at times. Throughout the sequence, the SNR by wavelet method is consistently higher than that of the moving average method, the grand mean (± SD) SNR were: SNR_{MA}: 12.06 ± 4.26; SNR_{DWT}: 19.15 ± 7.21; and SNR_{DWT}/SNR_{MA} =1.76 ± 0.89. SNR_{DWT} is superior to SNR_{MA}. Therefore, this example shows the benefit of wavelet regression versus a single scale moving average.

The improvement in signal-noise-ratio over a simple averaging approach is no coincidence: multiscale filtering using wavelets generally outperforms monoscale methods such as mean filtering and exponential smoothing for data that contain features localized at multiple scales [123] where data show jumps, spikes, and discontinuities as observed in LTHMP.

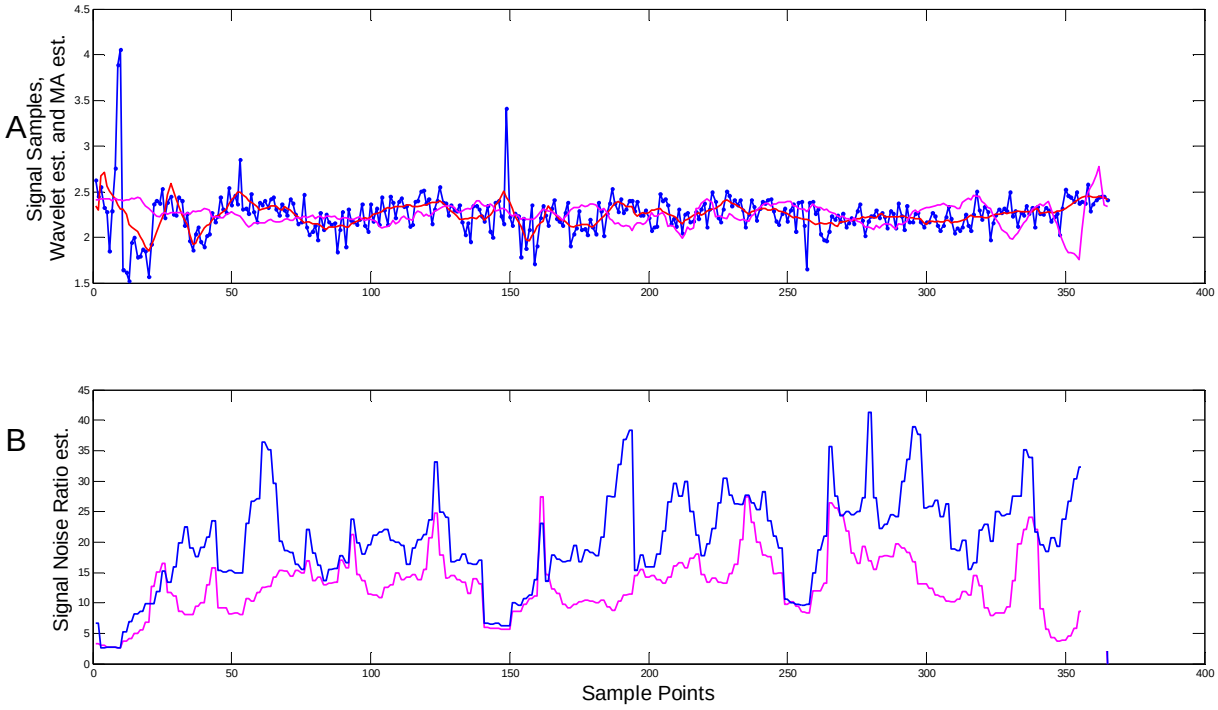

Figure 14. Signal Noise Ratio per DWT and MA

A: The upper figure shows the data sample points (connected with solid lines "—"); a wavelet filtered estimate "—", and a simple moving average filtered estimate "—".

B: The lower figure shows the signal noise ratio estimate for a wavelet method "—" and for a moving average method "—".

(Subject 5183)

Wavelet Multiscale Analysis

The Maximum Overlap Discrete Wavelet Transform (MODWT) [124], was used to simultaneously inspect the scale-dependent signal behaviors. MODWT is an un-decimated modification to DWT which is therefore time shift invariant, i.e., a translation in the signal will result in a translation of wavelet coefficients by the same amount [125]. Therefore, MODWT aligns wavelet coefficients at each time point with the original data index so one can simultaneously analyze localized signal variation with respect to scale and time, and the temporal relation to events.

This is particularly useful as an offline analysis, which can approximately estimate the event onsets corresponding to the event records, and in turn can estimate an appropriate event window size for detection classification.

Figure 15 illustrates an MOWDT analysis. The signal and time-dependent variations at each detail scale were examined simultaneously. Scale dependent signal variations stood out as the potential features related to the events. As highlighted in the boxes, prior to events, there was increase of variations at these fine scales, on top of a declining trend (shown as arrows). These behaviors are not easily seen in the single scale methods. From this illustration, the FEV_1 started to decline from the recent peaks (panel C) along with fine scale signal variations (panel B) at approximately 13, 8, and 17 days prior to their corresponding event records. Therefore, event onsets were approximated as having started 2~3 weeks prior to the corresponding event records.

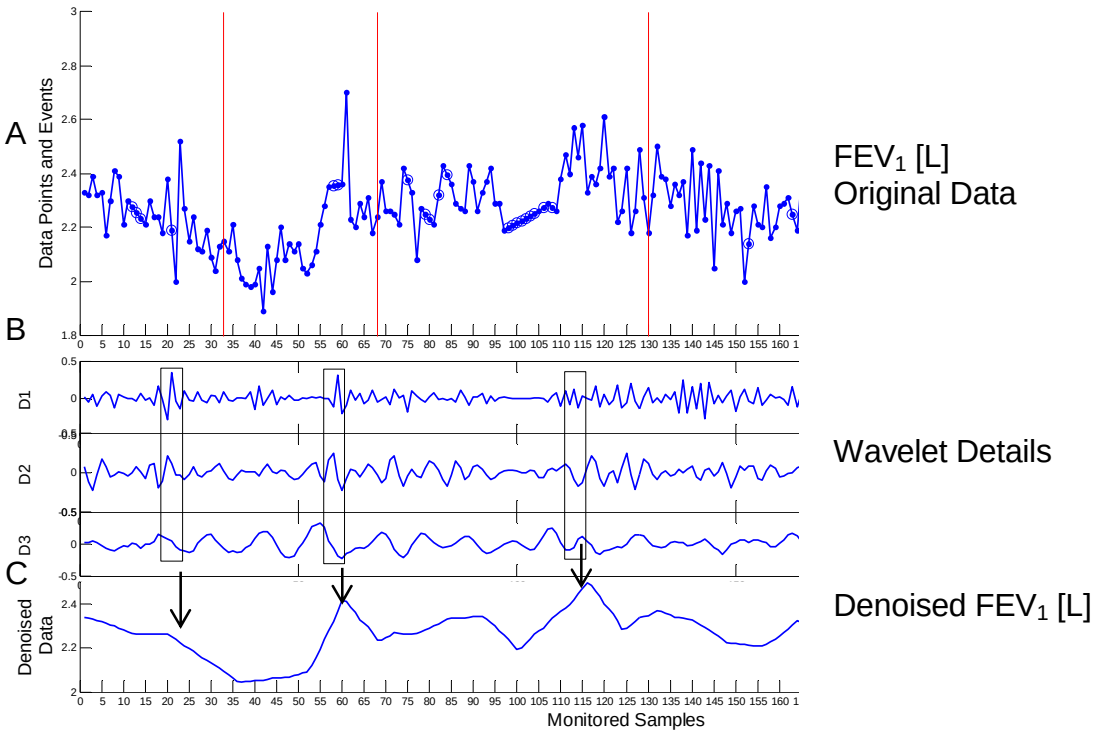

Figure 15. Estimating Event Onset and Defining Event Windows Using MODWT.

A: signal series with events, unfilled data points are linearly interpolated and shown with circles. B: Wavelet details plots (D1-D3); C: The denoised signal.

The red vertical line represents events records. The vertical boxes are approximate timing of event onsets.

(Subject X3033)

Computational burden of the DWT is in the order of $O(N)$ and $O(N\log_2 N)$ for MODWT, this is quite acceptable for practical use since they are at the same order of FFT (Fast Fourier Transform) ($O(N\log_2 N)$) or less.

Wavelet analyses were performed in MATLAB 7 computing environment with Wavelets Toolbox [126], user-developed Wavelab [127], and wavelet methods for time series [124].

Event Feature Extraction - A CUSUM Approach

Feature refers to a distinctive trait of the data. The data represent patients in a clinical setting who have unique disease history and unique reaction to lung transplants, showing wide range dynamics in their home monitoring data. Because individual response to disease process is complicated and not stereotypical between subjects or even between events in a single subject, the data pathway is unpredictable. Because no characteristic data patterns could be identified as event templates, event feature extraction warrants a data driven approach.

Through wavelet multiscale analysis of each and every spirometry variable data series of each subject, it was identified that unusual scale-dependent data variation amid a declining trend indicated events. The near term decline could be highlighted by subtracting a signal from the longer term trend as the reference, showing as “crossovers”. This idea is also consistent with the clinical guidelines to diagnose pulmonary diseases by comparing current spirometric readings to the previous readings to assess transplant status or disease progression [71].

Selection of Resolution Parameters

Through case analyses, a resolution scale of 2^3 days traced data closely. A scale of 2^6 days produced slow evolving trend that can serve as a moving reference. In general, crossover occurs more frequently when the scales are closer, with the extreme scale of 2^0 (raw data points) crossing over the reference constantly.

These choices of resolutions are similar to the parameters in the automated triage rules [17] in which moving average window sizes of a week is roughly 2^3 (8) days, and 2 months is roughly 2^6 (64) days, which also gave some assurance of these scale choices .

CUSUM of Signal Crossover for Event Detection

The CUmulative SUM (CUSUM) method, first described by Page in 1954 [128], is based on sequential monitoring of a cumulative performance measure over time. A sequential detection scheme is preferred in real time detection problems. The CUSUM methodology has emerged as a suitable method for physiological monitoring [129] and healthcare outcomes [130].

The idea of CUSUM is connected to a simple integration of signal over its target values to reveal a level shift over incremental samples. Such multiresolution signal and cross comparison has become a common theme in real time physiological data monitoring where measurement of sharp changes is contrasted with long-term trends (denoted in some literature as “offset”) [131]. This is suitable for on-line monitoring and event detection, regardless if changes are abrupt, rapid, or incremental. It is also a standard method in biomedical signal research to realize the event onset detection by defining the response onset as the point where the observed signal passes a certain threshold [132]. The CUSUM approach converts a fluctuating signal into a monotonic function, and prevents detection inconsistency, i.e., in alarm or out of alarm due to fluctuation of data around the threshold. Hence reliability is assured and the false alarm rate is controlled.

A CUSUM-based sequential event detector was therefore constructed. It is the running total of the differences between the signal and the reference, as previously described. The CUSUM is directional since only the one sided crossover indicating a declining pulmonary health is used in event detection.

Assume a process where the signal s_i is above $E(s_i)$, the CUSUM statistics is

$$C_k = \sum_{i=1}^{k} \{s_i - E(s_i)\}, \tag{6}$$

$$\mathrm{var}(C_k) = \sum_{i=1}^{k} \{\mathrm{var}(s_i^{\,u}) + \mathrm{var}(s_i^{\,v})\}$$

at sequential time point t_i, $i=1,2,\ldots k$. $E(s_i)$ is the reference and s_i is the signal as in the wavelet analysis. s_i^u and s_i^v are the variance terms, from the wavelet analysis at resolution u and v.

Alternatively, this may also be written in the form

$$C_k = max(C_{k-1} + s_i - E(s_i), 0) \quad (7)$$

$$C_0 = 0,$$

where $i=1,2,\ldots k$, for $s_i - E(s_i)$.

Note in this application, a signal s_i below reference $E(s_i)$ is of interest, as opposed to the conventions above, one only needs to rectify the signs accordingly.

In Figure 16, Panel A shows the data points $x[t_i]$, wavelet estimated signal and reference. Panel B shows the subtraction of signal from the reference and the crossovers; Panel C shows the cumulative crossover, with the one-sided sections highlighted.

Given the signal series $x=[x_{tx}, x_{tx+1}, \ldots, x_{tx+k}]$ and the reference series $y=[y_{tx}, y_{tx+1}, \ldots, y_{tx+k}]$ with crossover at tx , the one-sided cumulative sum S_k at time k is

$$(1)\ S_k = \sum_{i=1}^{k} d_i = \sum_{t_x+i} (x_i - y_i)_- ,$$

where d_i is the difference of x_i (signal) and y_i (reference), $i=1, .., k$, $(x_i - y_i)_-$ is the negative portion when $x_i < y_i$ at time instant i, and the positive portion of the crossover is back calculated as

$$(2)\ S_{-k'} = \sum_{i=0}^{-k'} d_i = \sum_{t_x+i} (x_i - y_i)_+ ,$$

where $i=0, .., -k'$, $(x_i - y_i)_+$ is the positive section when $x_i > y_i$ at time instant i.

For simplicity, CUSUM refers to $S_{sign(k)}$, where the sign indicates the one-sided CUSUM of interest.

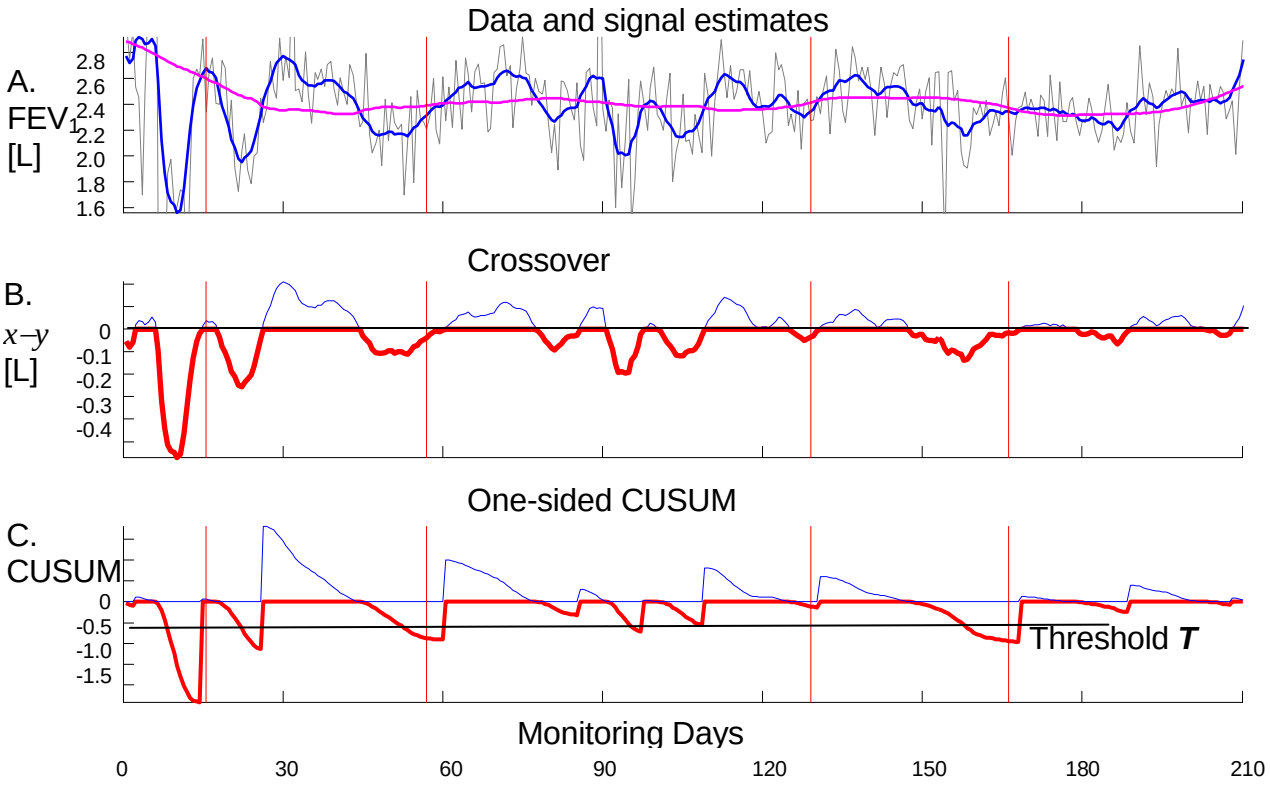

Figure 16. Illustration of signals, signal crossover and CUSUM of crossover for event detection.

From top to bottom: original data "—", signal(x "—") and reference(y "—") estimates (Panel A); crossover (one-sided "—", Panel B); cumulative crossover (one-sided "—", Panel C).

The initial CUSUM is set to zero. Once the signal crosses downward over the reference ($x_i < y_i$), the CUSUM starts to compute by summing up all the difference at each sample point i over the entire area under the curve (between 0 axis and $x_i - y_i$) as long as $x_i < y_i$. The CUSUM reaches its maximum when the signal upward crosses the reference ($x_i > y_i$) and is reset to zero. Once the next downward crossover occurs, the negative CUSUM starts computing again while the positive CUSUM is back calculated from the point of crossover. This way ensures

continuity at the crossover point. The one-sided CUSUM is further just referred as "CUSUM". When the CUSUM exceeds a pre-specified threshold value $\boldsymbol{T}$, the alarm sets off.

Because CUSUM is an integral form of the input series over time unit (day), its unit is [input unit-day], e.g., [Liter-day] for FEV_1. For brevity, we do not further specify the CUSUM unit.

Thus far, a single input vector $x \in R^1$ has been derived into a CUSUM vector $v \in R^1$, where $v[t_i] = f(x[t_i])$, at time t_i, i=1,2,3,...,n. Denote v as a "classifier" and the magnitude of each waveform v will be used in classification. The corresponding classification rule compares v to a predetermined threshold $\boldsymbol{T(T<0)}$ at each time instant, i. The event detection has thus been converted to a sequential detection problem.

The CUSUM classifier can be also viewed as if one starts scanning of signal at the current time backwards, until the CUSUM reaches a threshold, like a neighbor clustering for a spatial statistical analysis [130].

For detection using symptom converted score, the same CUSUM approach and computer code developed for the spirometry were applied. It is to be noted as opposed to spirometry, an increase in symptom is a sign of problem. By assigning a negative sign to the symptom series, the one-sided crossovers were used, consistent with those of spirometry.

The following illustrate event detection using CUSUM applied to symptom series.

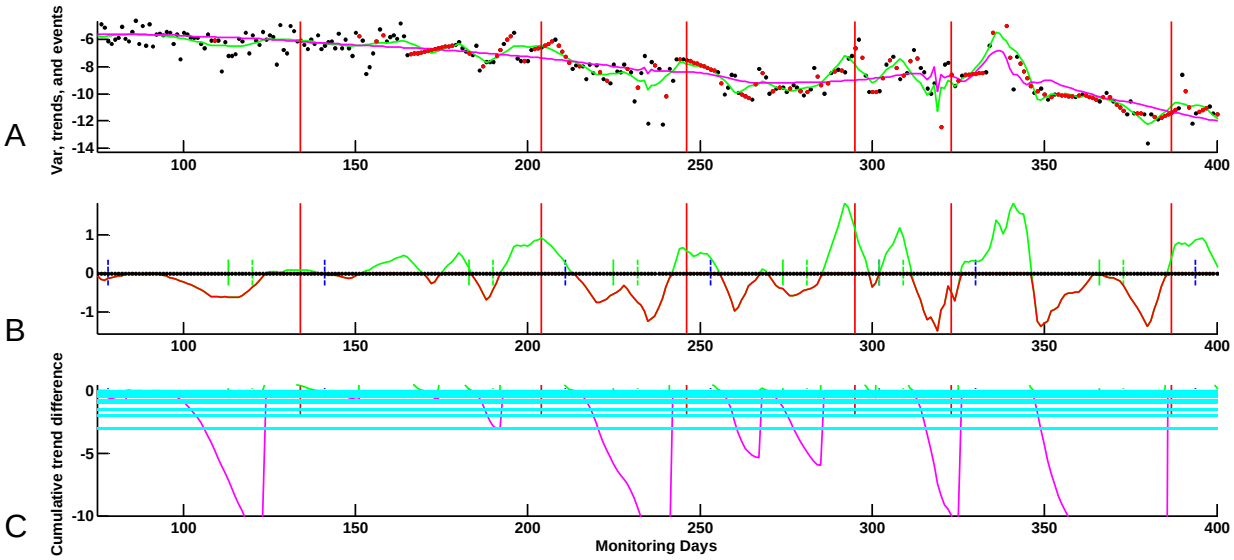

Figure 17. Illustration of symptom-based event detection.

Horizontal axis is the days of monitoring;

A : Data series is the symptom composite. The data points are dark dots (“.”), missing value replacement are red dots (“.”) with signal (“—”) and reference(“—”).

B : Difference between the signal and reference (crossovers) (one-sided); Small vertical lines marks 2,3 weeks prior to events and one week after events.

C : The one-sided cumulative crossover(“—”). The horizontal lines (“—”) are grid lines to determine if a threshold has been crossed.

Short vertical lines in B and C mark 2 weeks and 3 weeks prior to-, and 1 week post events.

(Subject X5153)

Summary of the Wavelet-based Detection

1. Select an input signal (e.g., FEV_1) time series vector $\boldsymbol{x}=x(t_i),\ i=1,2,\ldots,n$.

   a. Perform wavelet decompositions on ***x,*** at two resolution levels: 2^J days, J=3,6. Use Daubechies wavelets db3 for J=3 and db4 for J=6, respectively.

   b. Obtain wavelet coefficients ***w.*** Apply noise elimination with Universal Threshold formula ($T = \sigma\sqrt{2\log(n)}$) to reject or retain ***w***.

   c. Inverse wavelet transform from denoised ***w'*** resulting in smoothed signal vectors corresponding to high resolution ($\boldsymbol{y_1}$) and low resolution ($\boldsymbol{y_2}$).

2. Identify crossovers at points t_x where $y_1(t_x) - y_2(t_x) = d_{tx} = 0$. Subtract $\boldsymbol{y_1}$ from $\boldsymbol{y_2}$

3. Calculate cumulative sum (CUSUM S_k) :

   $S_k = \sum_{i=1}^{k} d_i$, for k points where d_i <0 corresponding to a decline.

4. Obtain a CUSUM time series vector $\boldsymbol{v} = v(t_i),\ i=1,2,\ldots,n$ to be classified.

   (MATLAB code in Appendix D)

Figure 18. Summary of the Wavelet-based Detection

Classification and Performance Evaluation

The performance metrics include the detection probability (sensitivity), the false alarm rate, and the timing of detection, following studies of continuous detection problem [133, 134, 135]. These metrics slightly differ from those of a binary outcome diagnostic test as in Table 6 below.

Performance Metrics

For binary classifications, a confusion matrix ("contingency table") below illustrates the classification of the testing samples. The true positive rate is TPR=TP/ N^+ and the false positive rate FPR= FP/N^-.

Table 6. Confusion matrix of a binary classifier

	Test Positive	Test Negative	Total Samples
Positive case	True Positive (TP)	False Negative (FN)	N^+
Negative case	False Positive (FP)	True Negative (TN)	N^-

Sensitivity

Sensitivity is the probability of true detection and is computed as the proportion of events that have been correctly identified by the detection algorithm. A true detection generates an alarm within the time frame associated with the event, i.e., an event window (also denoted as "signal occlusion period or SOP" in seizure detection literature).

Sensitivity is the same as the true positive rate (TPR)=TP/N^+. which can be normalized by the total monitoring duration - a constant value for the given data set, expressed as number/subject-year.

False Alarm Rate (FAR)

A false alarm is an alarm in absence of an event. If detection occurs outside of the event window, it is a false alarm (" false positive"). The false alarm rate (FAR) is defined as the number of false alarms divided by the total monitoring duration, expressed as number/subject-year.

$$FAR=FP/Duration$$

FAR is not the ratio of false alarms over all alarms as reported in some research.

True Alarm Rate (TAR)

Similar to the definition of FAR, one obtains TAR as $TAR=TP/Duration=Sensitivity\times N^{+}/Duration= Sensitivity \times const.$

Specificity

Specificity is defined as the probability of testing negative given a negative trial in a binary testing scenario (TN/N^{-}). Specificity is not well defined in sequential testing problems as there are infinite ways to segment data into individual trials. However, given FAR, one can estimate specificity per a procedure discussed in section "Approximating ROC". Since a high FAR indicates low specificity, specificity is inversely related to FAR.

Event Window for Classification

Classification is made by comparing the timing of detection to the corresponding event records. As previously illustrated in Figure 15, wavelet multiresolution data display and examination provided valuable insight to estimate event onset. As the event types are acute pulmonary-related in nature, on average, an event window was estimated to start typically within 2~3 weeks prior to the corresponding event records. Therefore, two clinically feasible window sizes were used: 1. An event window starts 2 weeks prior to event, and ends 1 week after event. 2. An event-window starts 3 weeks prior to event, and ends 1 week

after event. A fixed window size is only approximation, as quite often in medicine, the exact knowledge of event onset and its resolution is unknown, and sometimes only post hoc [136]. The remaining segments are defined as non-event related. The data series were thereby labeled with event status consistently for all input series prior to classification.

Classification

The following illustrates a representative event detection and classification scheme. Figure 19 shows the CUSUM sequences and their temporal relationship with the events (red vertical lines). Alarms are generated depending on whether the threshold line is crossed.

Most classifications are straightforward. By scanning the time series from left to right, the first CUSUM that exceeds a predetermined threshold will trigger detection. A detection occurs within the event window is a true detection. The true detections are marked by "✓" symbols (①-⑥ except for ④"7"). Subsequent detection of the same event by CUSUM ② is not recorded as a new detection and only one count of a true positive was recorded. Certain classifications may depend on the event window size (marked by "✓?" at ⑤ where a wider window would classify ⑤ as true vs. a narrower window would be false). Multiple false positive detections could happen consecutively and each would be counted as one false positive.

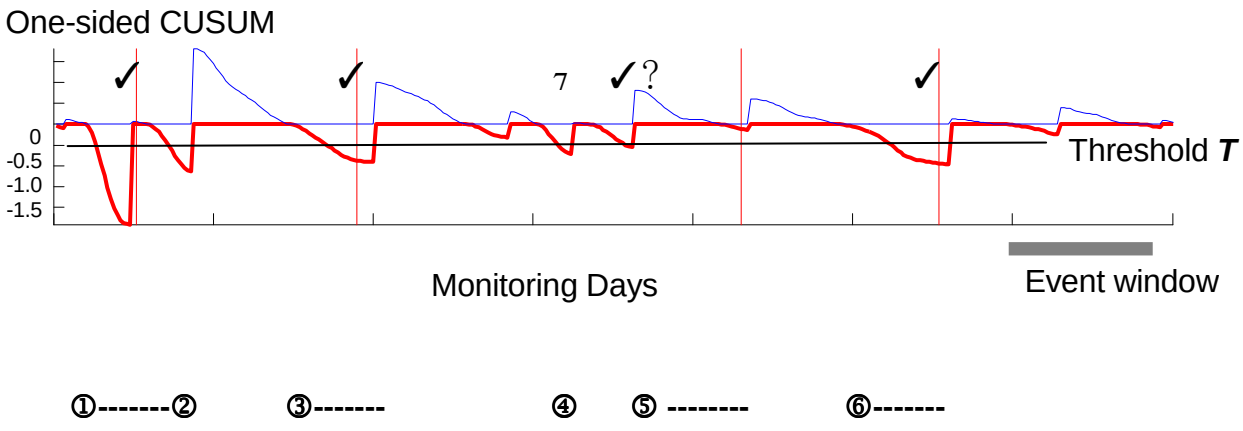

Figure 19. Illustration of event detection and classification through CUSUM.

To summarize, the CUSUM are classified per the following rule:

Classifying CUSUM

- Max(one-sided CUSUM) > |T| generates a detection.
- Detection within event window is a true positive and outside event window is a false positive.
- If a CUSUM detects an event at threshold |T|, it is detected at |T'|<|T|.
- There is at most one true detection for each event, which can be detected by a subsequent CUSUM.
- Each false positive was counted as a separate false positive.
- If a false CUSUM was more than 6 weeks wide, two false positives were counted.

Figure 20. Classifying CUSUM.

Operating Characteristics (OC)

The performance of a detection algorithm is characterized using an Operating Characteristics (OC) curve consisting of sensitivity versus false alarm rate. This is a well documented approach involving continuous monitoring [133,134,135] and seizure detection in EEG analysis [137,138]. OC curves are constructed by plotting the sensitivity and false alarm rate as a function of the detection threshold $\boldsymbol{T}$. A range of threshold levels were selected (shown as gridlines in plots) that would generate sensitivity between 0 and 1. The OC differs from the Receiver Operating Characteristics (ROC) in which the sensitivity and specificity are computed for discrete testing scenarios. The predicted best threshold $\boldsymbol{T}$ for future testing is selected by a tradeoff between A) maximizing sensitivity in detecting events, and B) minimizing the false alarm rate.

The plots were produced in Microsoft® Office EXCEL 2003, Chart Wizard XY scatter plot.

OC Weighting: Per-event and Per-subject

The OC performance was calculated according to the two ways of weighting: “per event” and “per subject”. In the per-event weighting, all events are equally weighted regardless from which subject. The sensitivity is the number of detected events divided by the total number of events. In the per subject weighting, the sensitivity is first calculated for that subject, and then averaged across all subjects. The per subject metric potentially avoids being overly influenced by a particular subject with many events.

Similarly, false alarm rate is calculated in two ways: the total number of false alarms divided by the total subject-years is denoted as “per event false alarm rate” (number/subject-year); the total number of false alarms divided by each subject-year, then averaging across all subjects is denoted as “per subject false alarm rate” (number/subject-year).

Approximating ROC

OC provides complete and unbiased performance metrics for the detection problem. However, ROC is the conventional metrics and this section describes the conversion of OC into ROC so one can also interpret the performance in the conventional terms.

Consider the monitoring dataset containing a total of N segments, of which N^+ are identified with events. Let this group be denoted by C_1, and let the group of the remaining $N^-=N - N^+$ not associated with events be denoted by C_2. For the mth monitoring variable, its detection performance is calculated by OC previously described. Recall an ROC for diagnosis is constructed by varying the threshold cut point.

Let $\hat{p}_j$ be the maximum of the CUSUM sequence v_j , $j \in Z$, and z the threshold cutoff. For each cut point z, the following quantities can be computed:

$$\begin{aligned} Y &= I(\hat{p}_j \geq z) \\ \bar{Y} &= I(\hat{p}_j < z) \end{aligned} \tag{8}$$

where $I(.)$ is an integer indicator function that the one-sided monotonic CUSUM sequence results in either $Y=1$ and $\bar{Y}=0$ or $Y=0$ and $\bar{Y}=1$.

$$\begin{aligned} \mathrm{POS(z)} &= \sum_{j \ni C_1} I(\hat{p}_j \geq z) \\ \mathrm{NEG(z)} &= \sum_{j \ni C_2} I(\hat{p}_j < z) \\ \mathrm{FP(z)} &= \sum_{j \ni C_2} I(\hat{p}_j \geq z) \\ \mathrm{FN(z)} &= \sum_{j \ni C_1} I(\hat{p}_j < z) \\ \mathrm{SENS(z)} &= \frac{\mathrm{POS(z)}}{N^+} \\ 1 - \mathrm{SPEC(z)} &= \frac{\mathrm{FP(z)}}{N^-} \end{aligned} \tag{9}$$

$\mathrm{POS}(z)$ = the number of true positive detection at threshold cut point z; $\mathrm{NEG}(z)$ = the number of true no detection at z; $\mathrm{FP}(z)$ = the number of false positive detection at z; $\mathrm{FN}(z)$ = the number of false negative detection at z; $\mathrm{SENS}(z)$ = sensitivity or the probability of true detection; $1 - \mathrm{SPEC}(z)$ = the probability of false detection.

As the relationship between false alarm rate (FAR) and $\mathrm{FP}(z)$ holds as

$$\mathrm{FAR} = \frac{\mathrm{FP}(z)}{Duration} = \frac{\mathrm{FP}(z)}{\overline{W}indow \cdot N} \tag{10}$$

for an average window size $\overline{W}indow$.

However, the reverse process does not hold since there are infinite ways to segment the data series, and each ROC depends on the segmentation. As previously explained, since one does not have the exact knowledge of event onset and resolution, one could not accurately and realistically perform segmentation upfront.

Therefore, $\mathrm{OC}(\mathrm{SENS}(z), \mathrm{FAR})$ for the sequential detection scheme is approximately equivalent to a conventional ROC $(\mathrm{SENS}(z), 1 - \mathrm{SPEC}(z))$. One can also discuss the performance in the conventional terms in ROC. The test statistic based on the area under the curve $A(z)$ [139]

$$\frac{A(z) - 0.5}{SE(A(z))} \tag{11}$$

which follows a standard Gaussian distribution with mean 0 and variance 1.

Timing of Detection

One important component in evaluating an event detection algorithm θ is the timing of the detection τ, which is defined as the earliest time t_a when CUSUM classifier S_k crosses the threshold $\boldsymbol{T}$, compared to the event time t_e. For a detectable event θ, the alarm time t_a is

$$t_a=\min\{k: |S_k|> |\boldsymbol{T}|\}. \quad (12)$$

$$\tau = E(t_e - t_a \mid \theta) \quad (13)$$

A time $\tau>0$ is "early warning" since the day of detection is prior to the day of corresponding event documentation. In the case of a late detection where $t_a > t_e$, τ <0 represents a delay in detection.

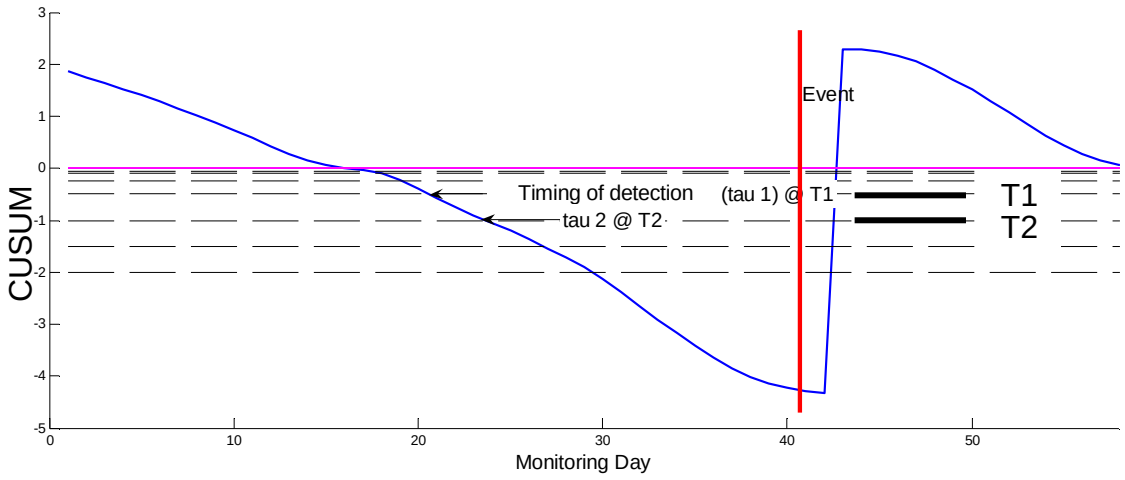

Figure 21. Timing of detection τ and early detection.

At threshold cutoffs |T1| < |T2|, the timing $\tau 1 > \tau 2$, i.e., sooner detection with lower cutoff values.

Results

In results, representative cases are first illustrated. In particular, a case with detailed event diary is used to provide connection between detection and clinical insight.

The overall performance including all subjects is summarized by the Operating Characteristic (OC) curves. Then the performance is further evaluated under the validation set.

Case Illustrations

A Case with Event Diary

A representative case with detailed event diaries is illustrated next. This is a single lung transplant male subject who had experienced two events within 250 days of follow-up. Clinical information is annotated chronologically with the FEV_1 home data.

Figure 22 shows how the detection works under the signal crossover method. The wavelet estimate of FEV_1 clearly indicates that about 2~3 weeks prior to each event, the data trended lower and crossed the reference, initiated a crossover. This crossover continued and the CUSUM increased and peaked about 4 and 5 days post the two events, respectively. The second detection would be classified as a false alarm due to lack of qualifying event documentation. This example contains insightful annotations to support our contention that the crossovers were indeed related to events and did not occur randomly.

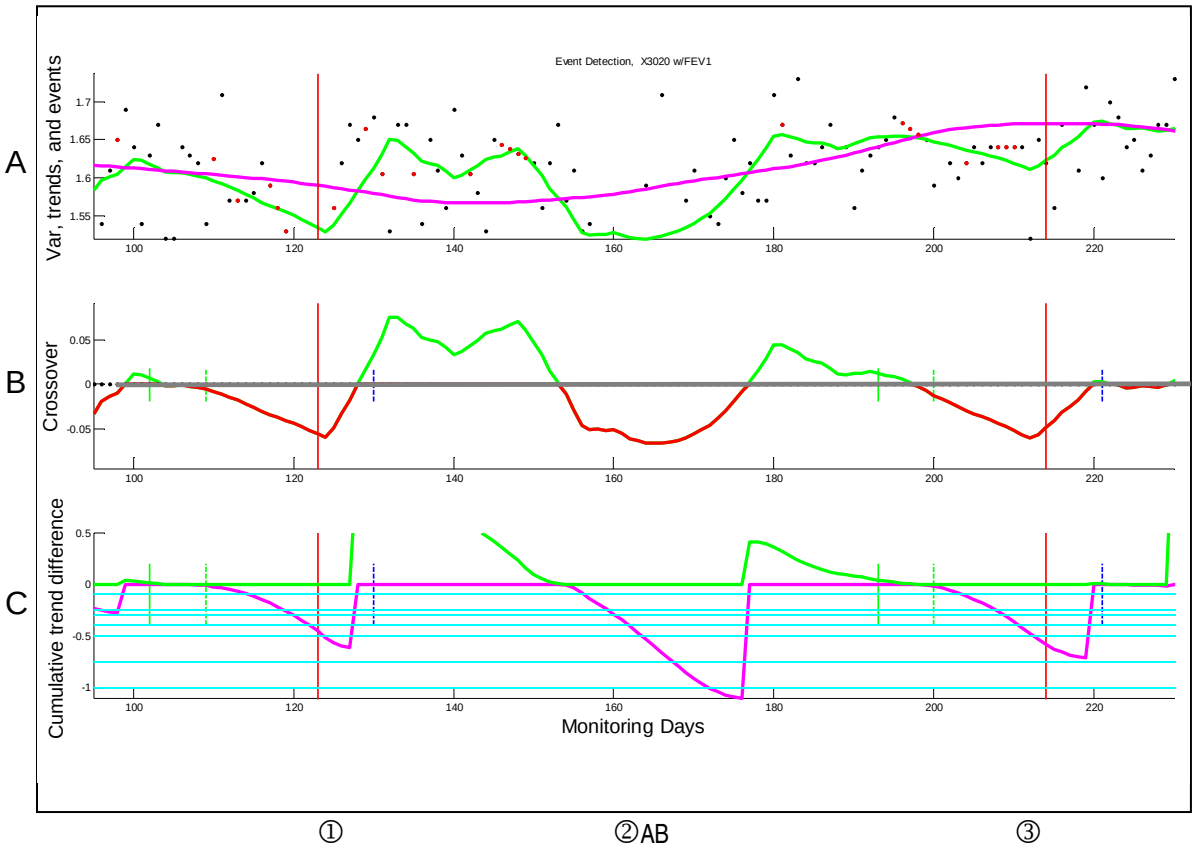

Figure 22. A case illustration of event detection based on FEV_1.

A: Data points, signals and events: original data points "." and interpolated data points ".", signal ("—") and reference ("—") [L];

B. Signal crosses over the reference ("—", one-sided);

C. CUSUM("—", one-sided). The long vertical lines indicate events.

B&C. Short vertical lines mark event window sizes of 2 weeks and 3 weeks prior to-, and 1 week post events to classify detections. The light blue grid lines in C "—" are grid lines which serve as thresholds.

(Subject X3020)

The event diary for this subject annotated the follow-up information which details the situation surrounding these events:

① Event 1. “Subject had bronchoscopy (day 123) -cultures showed moderate Pseudomnas growth” (event).

② “Subject called after finishing Ciprofloxacin which made him "tired”, he still has not recovered and has fatigue, slight shortness of breath (SOB), “sniffles" for 2 weeks. Has not seen his physician. Physician ordered labs including Prograf (Tacrolimus), CMV PCR, CBCP, electrolytes, and nasopharyngeal swab for RSV and viral culture. Subject was reminded to evaluate his clot and stopping warfarin. All orders faxed directly to subject so he may expedite process. “

A. Nurse note shows absence of event (day 161): “PCP (primary care physician) visit in emergency room - saw PCP for sore throat - no Tx (surgical procedure) or Rx (medical treatment)”.

B. Physician note shows absence of event (day 164): “FEV_1 trend down & increased symptom/severity; Had patient see PCP, exam negative. Physician ordered NSP for RSV and culture, labs. RSV negative, CX(chest x-ray) pending. Labs stable. FEV_1 level slightly low and dose adjusted. Subject felt he was more SOB in AM with cough, but by 12 noon he was less SOB and not coughing. If he did not get better or numbers did not go up, physician would ask that he came to clinic two weeks later for clinic and bronchoscopy”.

③ Event 2. Nurse note (day 212): “Upper respiratory sinusitis - amoxicillin provided and no other treatment.”

This case also indicate a “false” alarm, associated with ② per strict event definition in this study, which would still be clinically meaningful as an emergency visits was initiated.

Note: Prograf, PCR, CBCP, electrolytes, warfarin, Ciprofloxacin, etc., all refer to medical regimens.

The following illustrates event detection under considerable proportion of data omissions. As in Figure 23, there were 292 days of observations among 526 calendar days of monitoring (44% unfilled calendar) and the event detections. Data omission may affect detections in varying degrees, but in this case, the CUSUM sequences were correctly generated in 7 out of 7 events (100%) with few incorrect sequences.

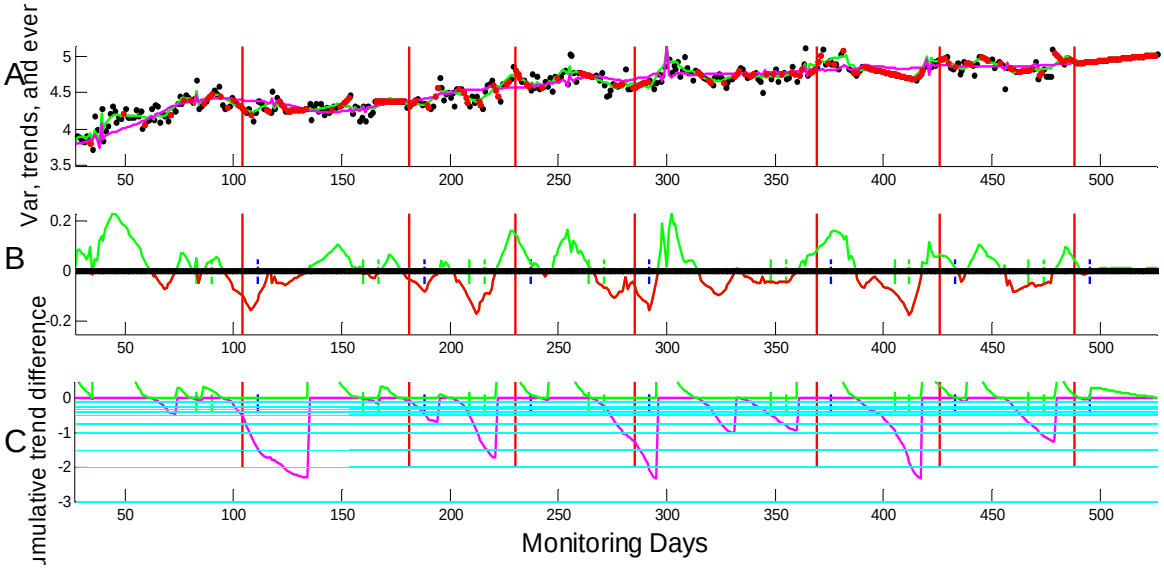

Figure 23. A case with data omissions in the learning set.

A: Data points, signals and events; B: Crossover; C: CUSUM (one-sided)

Labels are identical to previous plot. Omitted data points were interpolated as the red dots in the top plot.

(Subject X5145)

The following is a case in the validation set which also involved many omitted data points. There were 51 observations among 232 calendar days of monitoring (78% of calendar with no data). All CUSUM sequences were correctly generated in 6 out of 6 events (100% sensitivity).

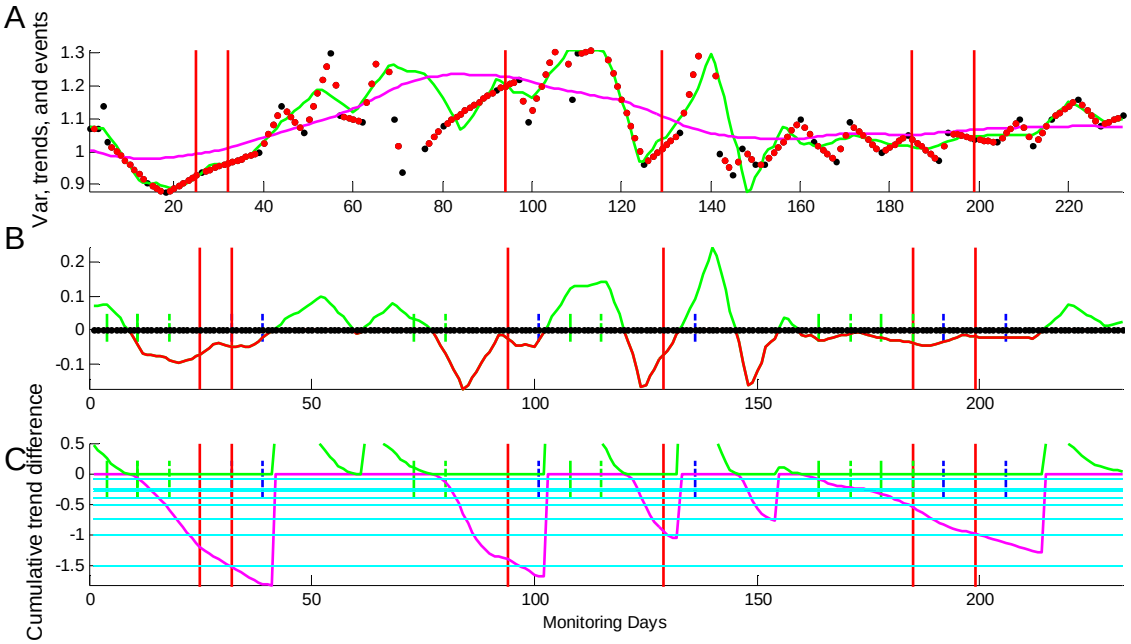

Figure 24. A case with data omissions in the validation set

A: Data points, signals and events; B: Crossover; C: CUSUM (one-sided)

Labels are identical to previous plot. Omitted data points were interpolated as the red dots in the top plot.

(Subject X3032)

Although these apparent successes might be due to the fact that this is a retrospective analysis where the data interpolation took advantage of data values made available at a later time, they also seem to suggest that data omissions that did not alter the trend would not severely impact the detections.

The Undetected Events

Certain events were undetected. The following describes the details of these undetected events using FEV_1 as the detection variable.

Of the 101 events in the learning set, 86 events would be detectable simply by adjusting the detection threshold ($|\boldsymbol{T}| \geq 0.1$). However, 15 events would not be detectable: in 10 events, the signal lines were above the reference lines (example shown in Figure 25) and crossover did not occur; 2 events were considered to contain critical data missing which were attributed to the events; in 1 event, both the raw data points and the signal line are ascending, which is exceptional to the basic assumption that events must be manifest with a decline in FEV_1; in 1 event, the data points jitter around a flat line which resulted in no CUSUM; in the last 1 event, there was a single observation of large decline in value which does not contribute to enough CUSUM for detection at $|\boldsymbol{T}| \geq 0.1$. Therefore, no triggering of crossover would be the main cause of detection failures. Cases are illustrated next in Figure 25 and Figure 26.

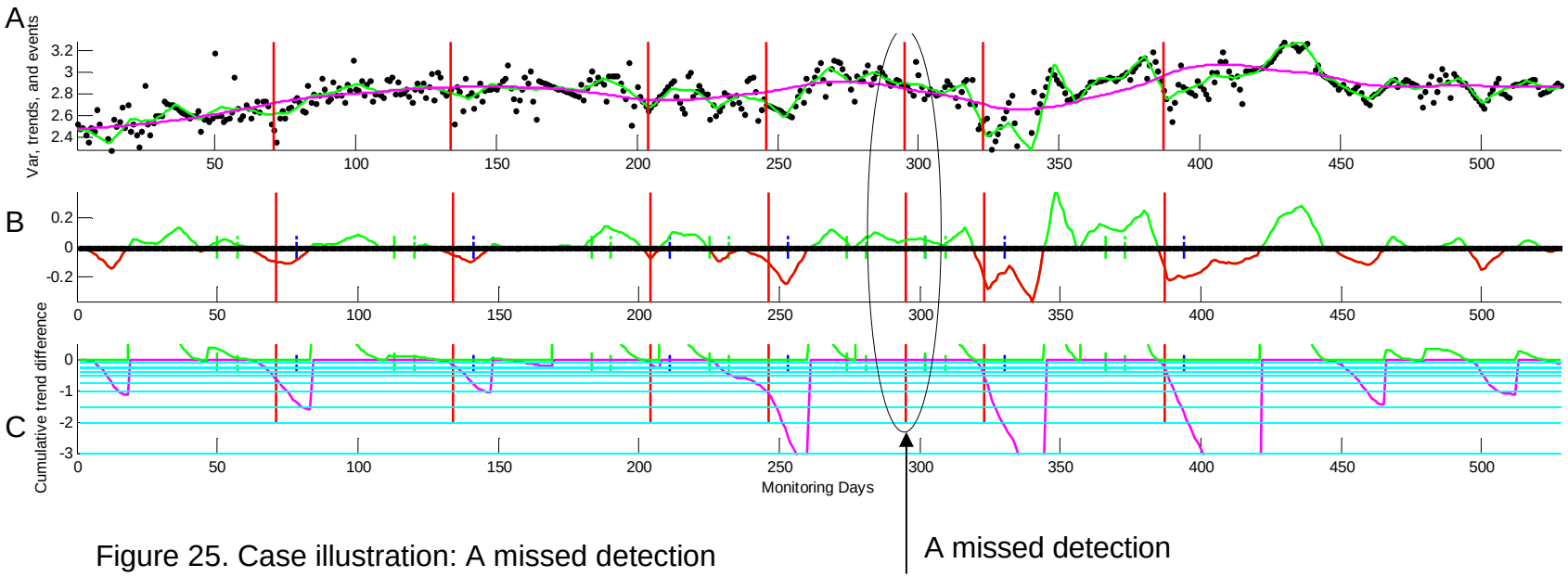

Figure 25. Case illustration: A missed detection

A: Data points, signals and events; B: Crossover; C: CUSUM (one-sided)

The signal line is above reference line, crossover never occurred, whereas all other 6 events were detected.

(Subject X5153)

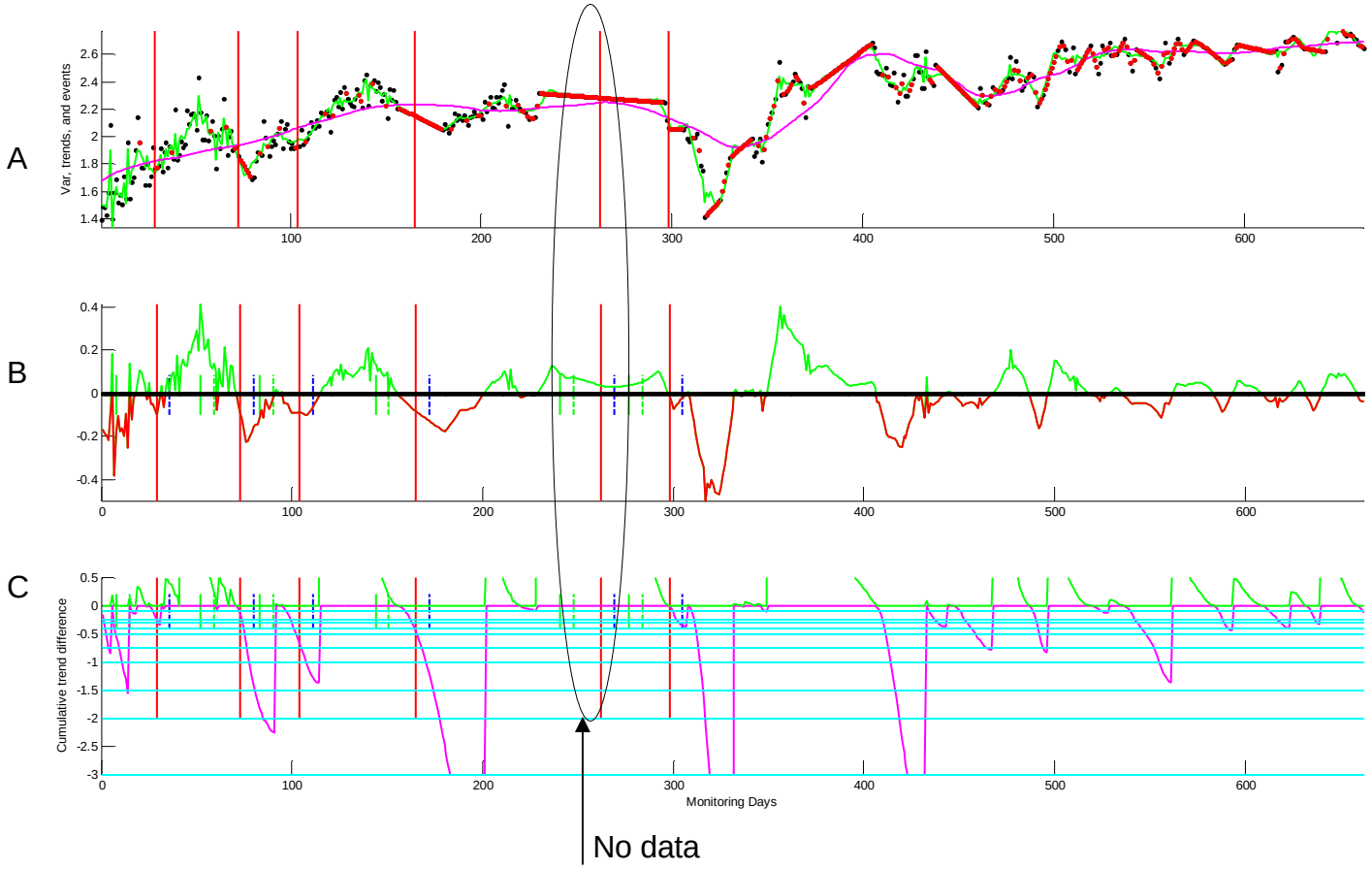

Figure 26. Case demo: A missed detection due to lack of data (missing data around the event).

A: Data points, signals and events; B: Crossover; C: CUSUM (one-sided). The interpolation to replace the missing data points, shows a flat line, did not generate crossover. (Subject 5170)

Performance Evaluation – Results from the Learning Set

Case illustrations show that detection algorithms work with varying degree of effectiveness for each subject. This section collectively reports the detection performance in all subjects involved in the learning set.

Operating Characteristics (OC)

To compare the overall performance of individual variables, one OC curve is constructed for each variable. A set of threshold values were selected for the cases in the learning set. The set of threshold value was independently chosen for each input variable, with some values overlap. To interpret the OC, for example, as the threshold $|T_{FEV1}|$ decreases from 2 to 0.1, the sensitivity and FAR increases, labeled as ①–⑥ in below and in Figure 27.

$\|T_{FEV1}\|$	2	1.5	1	0.5	0.25	0.1
Sensitivity	0.446	0.554	0.624	0.802	0.871	0.931
FAR	1.56	1.88	2.49	3.81	5.36	6.60
Label	①	②	③	④	⑤	⑥

Univariate OC's are plotted in the same plot, first by "per event" then by "per subject".

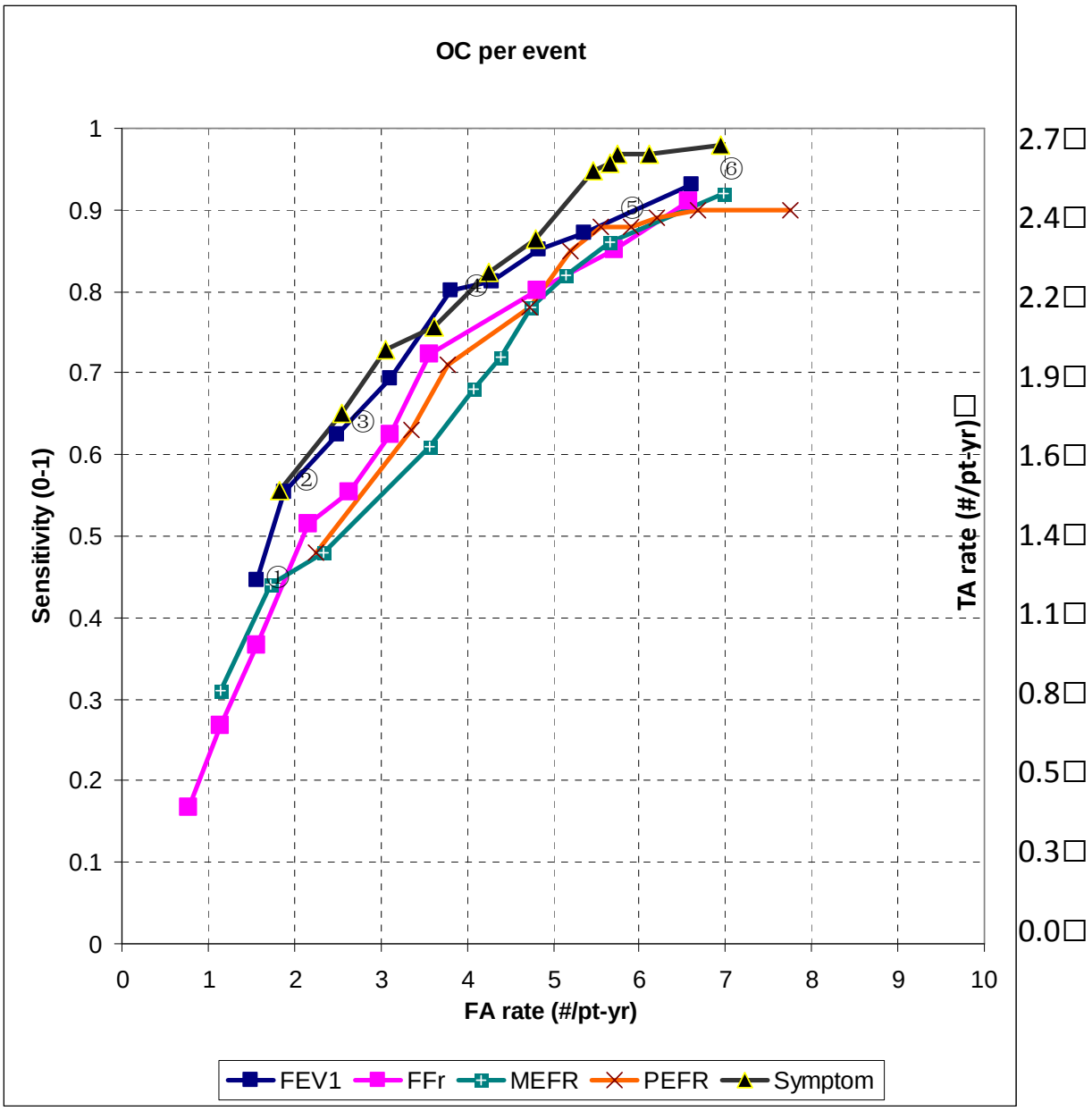

Figure 27. Operating Characteristics, per event.

Notes: Symptom: derived symptom index (page 37).

Axis on the right: *True Alarm Rate (TAR),* where $TAR=TP/Duration=Sensitivity\times N^{+}/Duration = Sensitivity\times(101/37.3)= Sensitivity\times 2.71$. One can easily convert the scale of sensitivity to *TAR*, e.g., an 80% sensitivity = 2.2 *TA/pt-yr*.

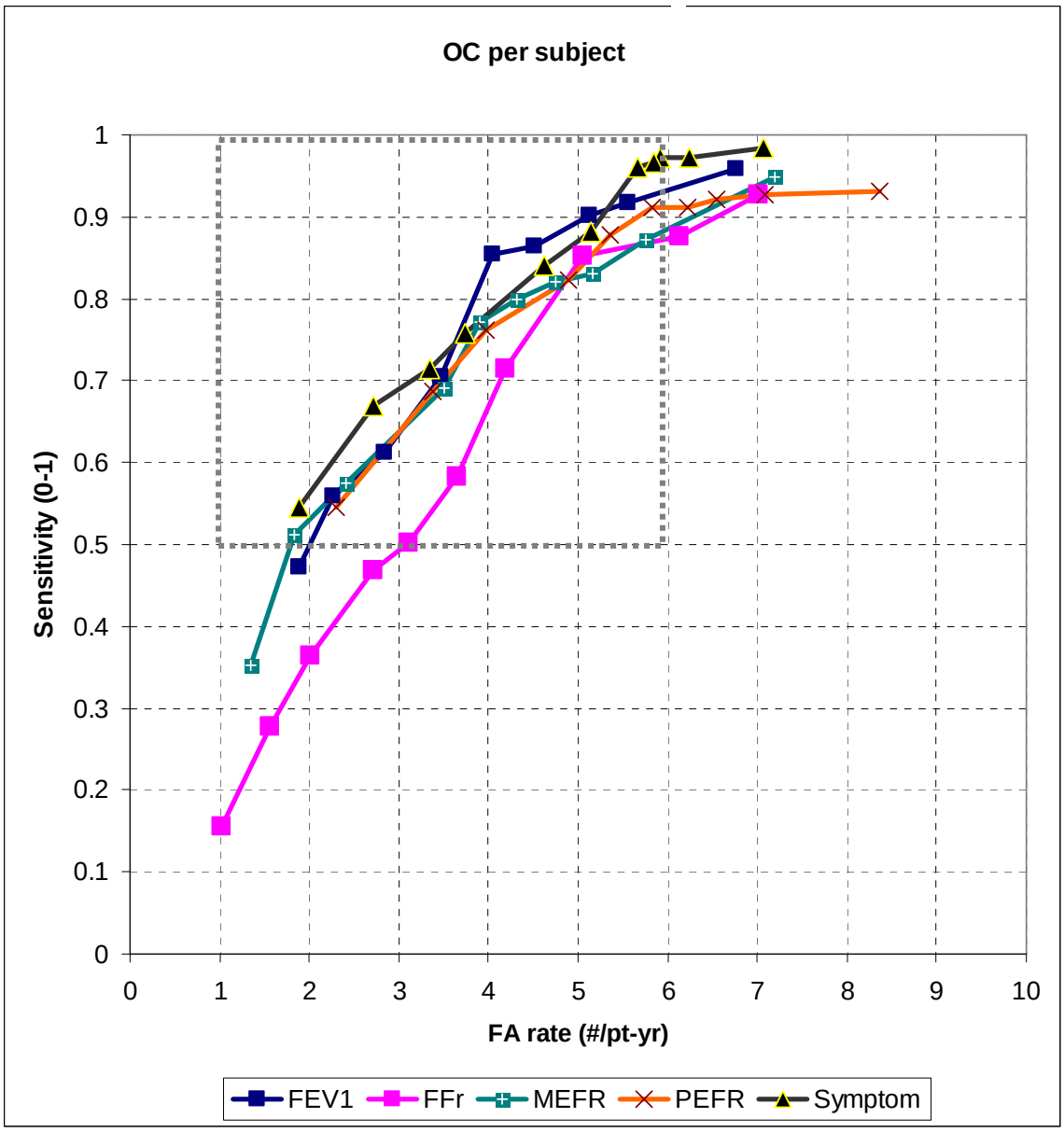

Figure 28. Operating Characteristics, per subject.

A partial area region is illustrated by the dashed lines which confine region: Sensitivity [0.5-1] and False Alarm Rate [1-6]/subject-year, where the partial "under the curve" (pAUC) is the proportion within the region.

Table 7. OC performance, per event

	FEV_1		FEV_1/FVC		MEFR		PEFR		Symptom	
$\|T_k\|$	S	FAR	S	FAR	S	FAR	S	FAR	S	FAR
0.05	--	--	0.911	6.57	--	--	--	--	--	--
0.1	0.931	6.60	0.852	5.71	0.921	6.97	0.905	7.75	0.979	6.95
0.25	0.871	5.36	0.801	4.81	0.862	5.66	0.900	6.68	0.969	6.12
0.3	0.851	4.83	0.723	3.57	0.823	5.15	0.891	6.20	0.969	5.74
0.4	0.812	4.29	0.624	3.11	0.782	4.75	0.882	5.90	0.958	5.66
0.5	0.802	3.81	0.554	2.63	0.721	4.40	0.882	5.55	0.948	5.47
0.75	0.693	3.11	0.515	2.17	0.681	4.08	0.853	5.20	0.865	4.80
1	0.624	2.49	0.366	1.56	0.614	3.57	0.781	4.72	0.823	4.26
1.5	0.554	1.88	0.267	1.15	0.485	2.33	0.714	3.78	0.757	3.62
2	0.446	1.56	0.168	0.78	0.444	1.72	0.635	3.35	0.729	3.06
3	--	--	0.139	0.46	0.317	1.15	0.481	2.25	0.650	2.55
3.5	--	--	--	--	--	--	--	--	0.555	1.82
pAUC	0.433		0.344		0.326		0.280		0.471	

Notes: S: Sensitivity (%event); FAR: False Alarm Rate(#/subject- year); "--" not evaluated. pAUC(0-1): partial area "under the curve".

Table 8. OC performance, per subject

	FEV_1		FEV_1/FVC		MEFR		PEFR		Symptom	
$\|T_k\|$	S	FAR	S	FAR	S	FAR	S	FAR	S	FAR
0.05	--	--	0.928	7.01	--	--	--	--	--	--
0.1	0.958	6.77	0.876	6.14	0.948	7.19	0.931	8.36	0.985	7.06
0.25	0.918	5.56	0.853	5.05	0.873	5.75	0.928	7.09	0.972	6.24
0.3	0.901	5.12	0.715	4.19	0.831	5.16	0.921	6.54	0.972	5.91
0.4	0.865	4.51	0.582	3.66	0.821	4.74	0.911	6.22	0.967	5.85
0.5	0.854	4.05	0.501	3.12	0.800	4.32	0.911	5.82	0.961	5.67
0.75	0.705	3.48	0.468	2.71	0.772	3.90	0.878	5.36	0.882	5.14
1	0.612	2.85	0.365	2.02	0.691	3.50	0.822	4.90	0.840	4.62
1.5	0.559	2.26	0.277	1.57	0.575	2.41	0.761	3.97	0.757	3.74
2	0.473	1.89	0.155	1.03	0.512	1.84	0.687	3.38	0.714	3.35
3	--	--	0.124	0.69	0.353	1.34	0.545	2.30	0.670	2.71
3.5	--	--	--	--	--	--	--	--	0.545	1.89
pAUC	0.434		0.250		0.379		0.381		0.448	

Notes: S: Sensitivity (%event); FAR: False Alarm Rate(#/subject- year); "--" not evaluated. pAUC(0-1): partial area "under the curve".

In OC performance, symptom achieves higher per event sensitivity at higher false alarm rate (FAR) (e.g., sensitivity 94.8%, FAR 5.47 /subject-year); at lower FAR, FEV_1 achieves higher sensitivity (e.g., sensitivity 80.2%, FAR 3.81 /subject-year). Similarly, symptom achieves higher per subject sensitivity at higher false alarm rate (FAR) (e.g., sensitivity 96.1%, FAR 5.67/subject-year); at lower FAR, FEV_1 achieves higher sensitivity (e.g., sensitivity 85.4%, 4.05 /subject-year).

With this set of data, the OC curves overlap. By constructing a square box region confining sensitivity [0.5-1] and false alarm rate [1-6]/subject-year as the focus area of interest, the proportion of the partial area under the curve to the total area of the region was defined as pAUC (0-1). The idea of pAUC is well documented in the ROC literature [140]. Two adjacent OC data points are connected with a straight line and the area is summed up using a trapezoidal formula. In the per event metrics, both FEV_1 (pAUC=0.433) and symptom (pAUC=0.471) were superior to the remaining variables (max pAUC=0.344); and in the per subject metrics, both FEV_1 (pAUC=0.434) and symptom (pAUC=0.448) were superior to the remaining variables (max pAUC=0.381).

ROC Approximation

The OC can be approximately converted to ROC, the conventional metrics for binary trials, as previously described. The entire area under the curve (AUC (0-1)) of ROC is interpreted as the probability of true positive detection [141, 142]. Figure 29 shows the ROC curves, which is by definition, per event metrics

The ROC result is consistent with the OC: Overall ranking per AUC was symptom (0.8360), FEV_1 (0.8194), MEFR (0.8068), PEFR (0.7955) and FEV_1/FVC ratio (0.7875). Symptom and FEV_1 remain the best two best variables for event detection. As the results were from a limited sample of 28 subjects, the probability of true positive statistics against a random guess (50% chance) were calculated per formulae in [139] and all p-values <0.001. Therefore, all of the AUC estimates suggest discriminating power to separate events from non-event status compared

to a random guess. In particular, FEV_1 observed sensitivity=85.5% at specificity=72.1% and symptom observed sensitivity=84.0% at specificity=68.2%. These OC curves do not exactly overlap. Symptom achieves higher sensitivity at lower specificity than FEV_1, e.g., sensitivity 96.1% at specificity 60.9%, suggesting symptom detects more event at the cost of lower sensitivity.

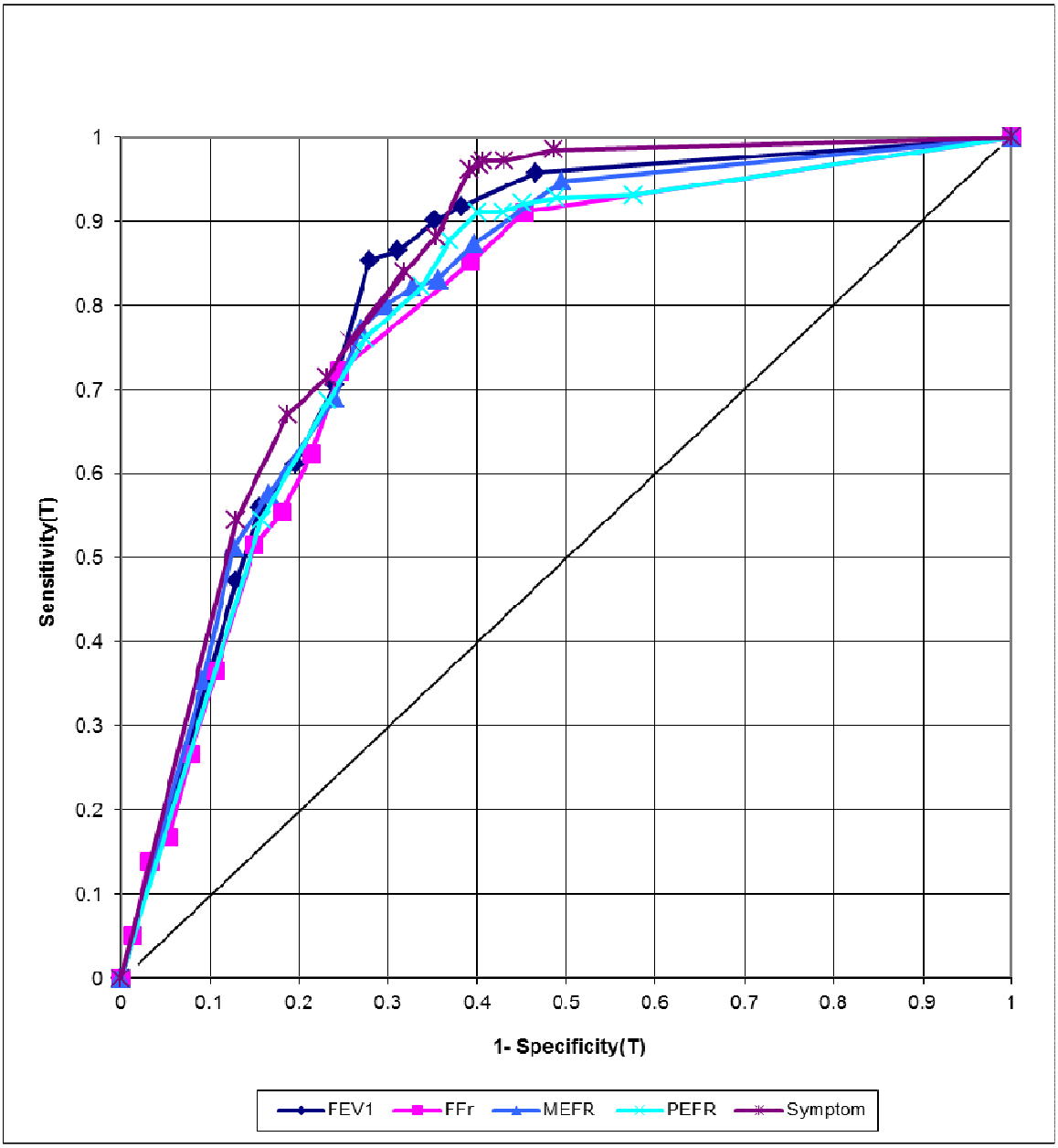

Figure 29. ROC approximation.

Table 9. ROC approximation.

	FEV_1		FEV_1/FVC		MEFR		PEFR		Symptom	
$\|T_k\|$	S	1-SPEC	S	1-SPEC	S	1-SPEC	S	1-SPEC	S	1-SPEC
0.05	--	--	0.928	0.453	--	--	--	--	--	--
0.1	0.958	0.466	0.876	0.393	0.948	0.495	0.931	0.576	0.985	0.486
0.25	0.918	0.383	0.853	0.246	0.873	0.396	0.928	0.489	0.972	0.430
0.3	0.901	0.353	0.715	0.214	0.831	0.356	0.921	0.451	0.972	0.407
0.4	0.865	0.311	0.582	0.181	0.821	0.327	0.911	0.429	0.967	0.403
0.5	0.854	0.279	0.501	0.150	0.800	0.298	0.911	0.401	0.961	0.391
0.75	0.705	0.240	0.468	0.107	0.772	0.269	0.878	0.369	0.882	0.354
1	0.612	0.196	0.365	0.079	0.691	0.241	0.822	0.338	0.840	0.318
1.5	0.559	0.156	0.277	0.054	0.575	0.166	0.761	0.274	0.757	0.258
2	0.473	0.130	0.155	0.032	0.512	0.127	0.687	0.233	0.714	0.231
3	--	--	0.124	0.013	0.353	0.092	0.545	0.158	0.670	0.187
3.5	--	--	--	--	--	--	--	--	0.545	0.130
Total AUC	0.8194		0.7875		0.8068		0.7955		0.8360	

Timing of Detection

Early warning time is measured by timing of detection, τ. τ is estimated by the time when threshold was first crossed as the earliest possible detection time and not truncated by the fixed window size of 3 weeks. The summary statistics of τ were computed for FEV_1 and symptom as below.

Table 10. Timing of detection τ estimate (days) in the learning set

Variable	Mean±SD	Median	Range	Threshold	Events Evaluated
				$\|\mathbf{T}_k\|$	N
FEV_1	10.8±9.6	12.6	45 ~ –13*	0.5	81
Symptom	10.0±13.5	10.2	48 ~ –31*	1	75

Note: *τ < 0 indicated a delayed detection compared to the corresponding event record.

On average, timing of early detection by FEV_1 and symptom was 10.8 and 10.0 days, respectively. FEV_1 typically leads symptom in detection timing. The mode of detection was at τ=0, coinciding with timing of the corresponding event records.

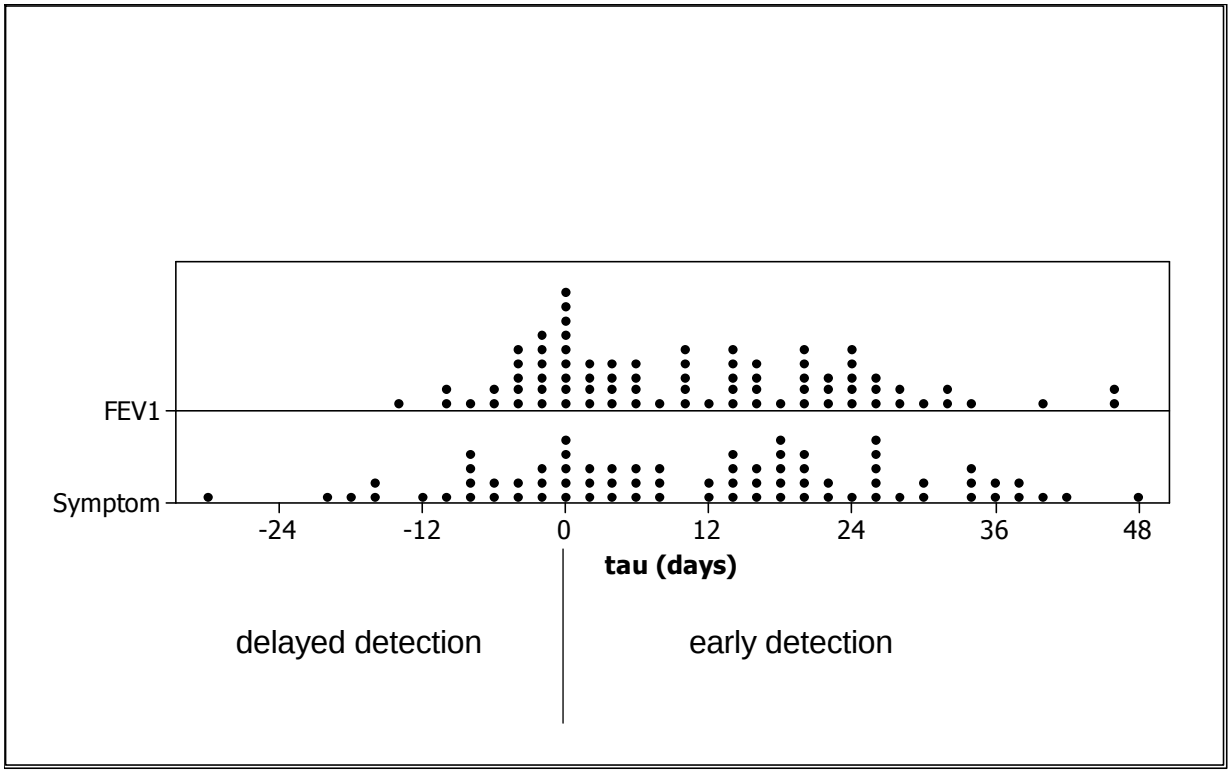

Figure 30. The distribution of timing of detection τ (days).

Performance Evaluation – Results from the Validation Set

The same detection methods and threshold levels evaluated on the learning set were applied in the validation set. OC for the validation set is shown as follows.

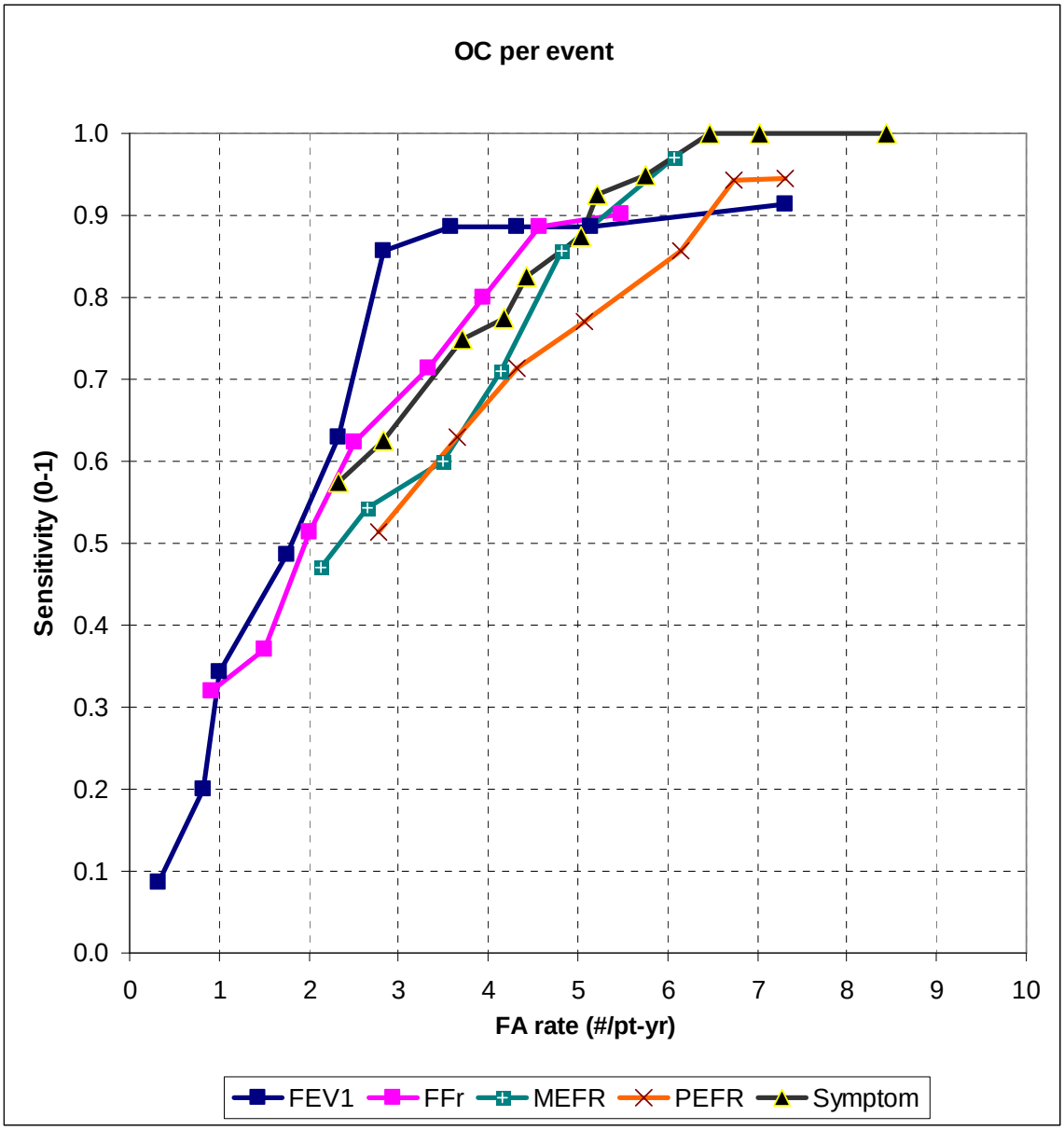

Figure 31. The Operating Characteristics of individual variables per event.

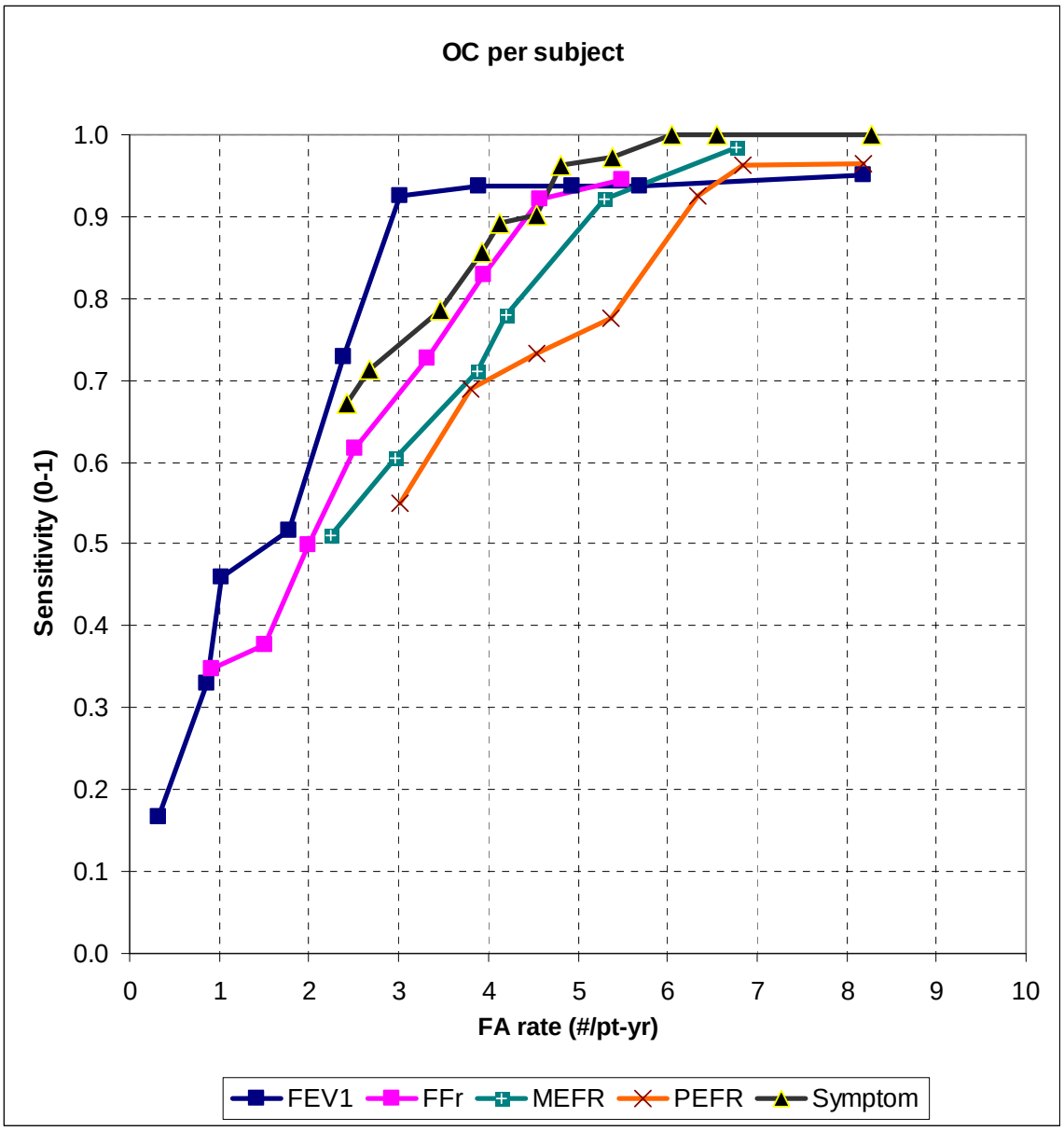

Figure 32. The Operating Characteristics of individual variables per subject.

Table 11. OC performance, per event

	FEV_1		FEV_1/FVC		MEFR		PEFR		Symptom	
$\|T_k\|$	S	FAR	S	FAR	S	FAR	S	FAR	S	FAR
0.05	--	--	0.901	5.49	--	--	--	--	--	--
0.1	0.914	7.32	0.886	4.57	--	--	0.945	7.32	1.000	8.44
0.25	0.886	5.15	0.623	2.51	0.971	6.07	0.943	6.73	1.000	7.02
0.3	0.886	4.32	0.514	2.00	--	--	0.857	6.15	1.000	6.47
0.4	0.886	3.58	0.371	1.50	--	--	0.771	5.07	0.950	5.76
0.5	0.857	2.83	0.320	0.91	0.857	4.82	0.714	4.32	0.925	5.21
0.75	0.629	2.33	--	--	0.710	4.14	0.629	3.66	0.875	5.03
1	0.486	1.75	--	--	0.600	3.49	0.513	2.78	0.825	4.42
1.5	0.343	1.00	--	--	0.543	2.66	--	--	0.775	4.18
2	0.200	0.83	--	--	0.470	2.14	--	--	0.750	3.71
3	0.086	0.33	--	--	--	--	--	--	0.625	2.84
3.5	--	--	--	--	--	--	--	--	0.575	2.33
pAUC	0.547		0.445		0.333		0.257		0.424	

Notes: S: Sensitivity (%event); FAR: False Alarm Rate (#/subject-year); "--" some previous threshold levels were not evaluated as they fell away from the pAUC. pAUC: partial area "under the curve".

Table 12. OC performance, per subject

	FEV_1		FEV_1/FVC		MEFR		PEFR		Symptom	
T_k	S	FAR	S	FAR	S	FAR	S	FAR	S	FAR
0.05	--	-	0.945	5.57	--	--	--	--	--	--
0.1	0.951	8.19	0.921	4.70	--	--	0.965	8.19	1.000	8.27
0.25	0.938	5.70	0.617	2.54	0.985	6.76	0.962	6.84	1.000	6.56
0.3	0.938	4.93	0.499	2.07	--	--	0.925	6.33	1.000	6.05
0.4	0.938	3.90	0.378	1.38	--	--	0.777	5.36	0.972	5.39
0.5	0.925	3.01	0.347	0.91	0.921	5.29	0.732	4.55	0.962	4.82
0.75	0.728	2.38	--	--	0.780	4.21	0.690	3.81	0.902	4.54
1	0.516	1.78	--	--	0.711	3.87	0.551	3.01	0.892	4.13
1.5	0.460	1.02	--	--	0.605	2.97	--	--	0.856	3.94
2	0.331	0.87	--	--	0.51	2.25	--	--	0.785	3.46
3	0.167	0.33	--	--	--	--	--	--	0.713	2.67
3.5	--	--	--	--	--	--	--	--	0.671	2.43
pAUC	0.629		0.468		0.384		0.270		0.549	

Notes: S: Sensitivity (%event); FAR: False Alarm Rate (#/subject-year); "--" not evaluated. pAUC: partial area "under the curve".

In OC performance, the highest per event sensitivity was achieved by symptom (e.g., Sensitivity 95.0%, FAR 5.75 /subject-year); at fixed FAR, the highest sensitivity was by FEV_1 (e.g., Sensitivity 85.7%, FAR 2.83 /subject-year). pAUC of FEV_1 and symptom was 0.547 and 0.424, respectively.

Per subject, the highest sensitivity was achieved by symptom (e.g., Sensitivity 97.2%, FAR 5.39/subject-year); at fixed FAR, the highest sensitivity was by FEV_1 (e.g., Sensitivity 92.5% , 3.01 /subject-year). pAUC of FEV_1 and symptom was 0.629 and 0.549, respectively.

The OC results from the validation set are generally consistent with the learning set, suggesting the detection method, designed based on the learning set, can be successfully implemented with new, previously unseen data. It confirms that FEV_1 and symptom show consistent performance and are the preferred variables for event detection.

ROC Approximation

Figure 33 shows the ROC curves.

The ROC result is consistent to the OC: Overall ranking per total AUC was FEV_1 (0.8279), symptom (0.8180), FEV_1/FVC ratio (0.8130), MEFR (0.8035), and PEFR (0.7695). The probability of true positive statistics against a random guess (50% chance) per formulae in [139] calculated all p-values <0.001. Therefore, all of the AUC estimates suggest discriminating power to separate events from non-event status compared to a random guess. In particular, FEV_1 observed sensitivity=85.7% at specificity=80.5% and symptom observed sensitivity=82.5% at specificity=69.5%.

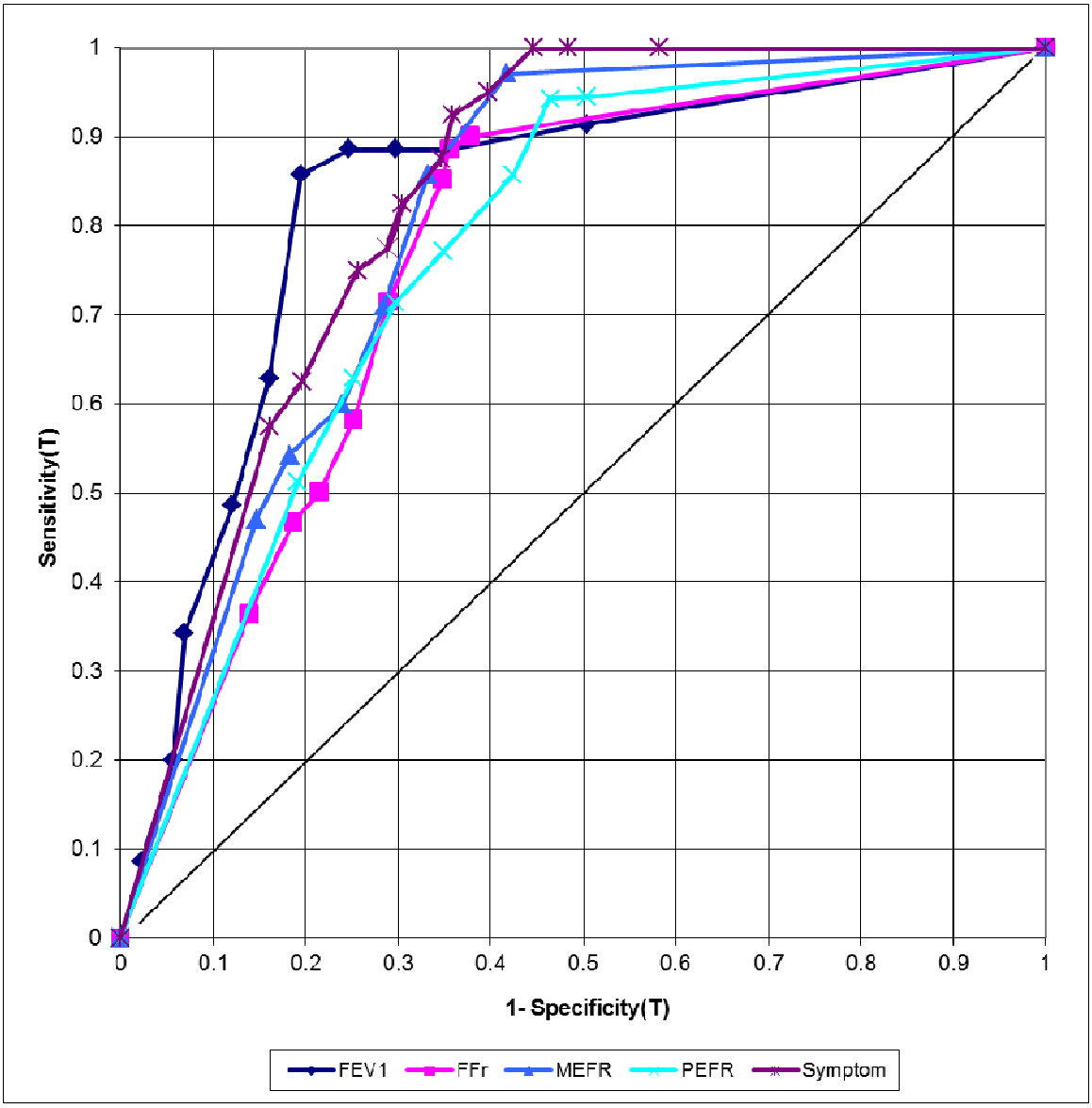

Figure 33. ROC approximation.

Table 13. ROC approximation.

	FEV_1		FEV_1/FVC		MEFR		PEFR		Symptom	
$\|T_k\|$	S	1-SPEC	S	1-SPEC	S	1-SPEC	S	1-SPEC	S	1-SPEC
0.05	--	--	0.901	0.378	--	--	--	--	--	--
0.1	0.914	0.504	0.886	0.355	--	--	0.945	0.504	1	0.582
0.25	0.886	0.355	0.623	0.252	0.971	0.418	0.943	0.464	1	0.484
0.3	0.886	0.298	0.514	0.215	--	--	0.857	0.424	1	0.446
0.4	0.886	0.247	0.371	0.187	--	--	0.771	0.349	0.950	0.397
0.5	0.857	0.195	0.320	0.139	0.857	0.332	0.714	0.298	0.925	0.359
0.75	0.629	0.161	--	--	0.710	0.285	0.629	0.252	0.875	0.347
1	0.486	0.121	--	--	0.600	0.24	0.513	0.192	0.825	0.305
1.5	0.343	0.069	--	--	0.543	0.183	--	--	0.775	0.288
2	0.200	0.057	--	--	0.470	0.147	--	--	0.750	0.256
3	0.086	0.023	--	--	--	--	--	--	0.625	0.196
3.5	--	--	--	--	--	--	--	--	0.575	0.161
Total AUC	0.8279		0.8130		0.8035		0.7695		0.8180	

Timing of detection

On average, the timing of early detection by FEV_1 and Symptom was 7.7 and 6.6 days, respectively, summarized Table 14 in and depicted in Figure 34.

Table 14. Timing of detection τ Estimate (days), validation set

Input Variable	Mean±SD	Median	Range	Threshold	Events Evaluated
				$\|T_k\|$	N
FEV_1	7.7±6.6	9.0	29~–18*	0.5	31
Symptom	6.6±6.7	4.8	26~ – 8*	1	25

Note: * $\tau < 0$ indicated a delayed detection compared to the corresponding event record.

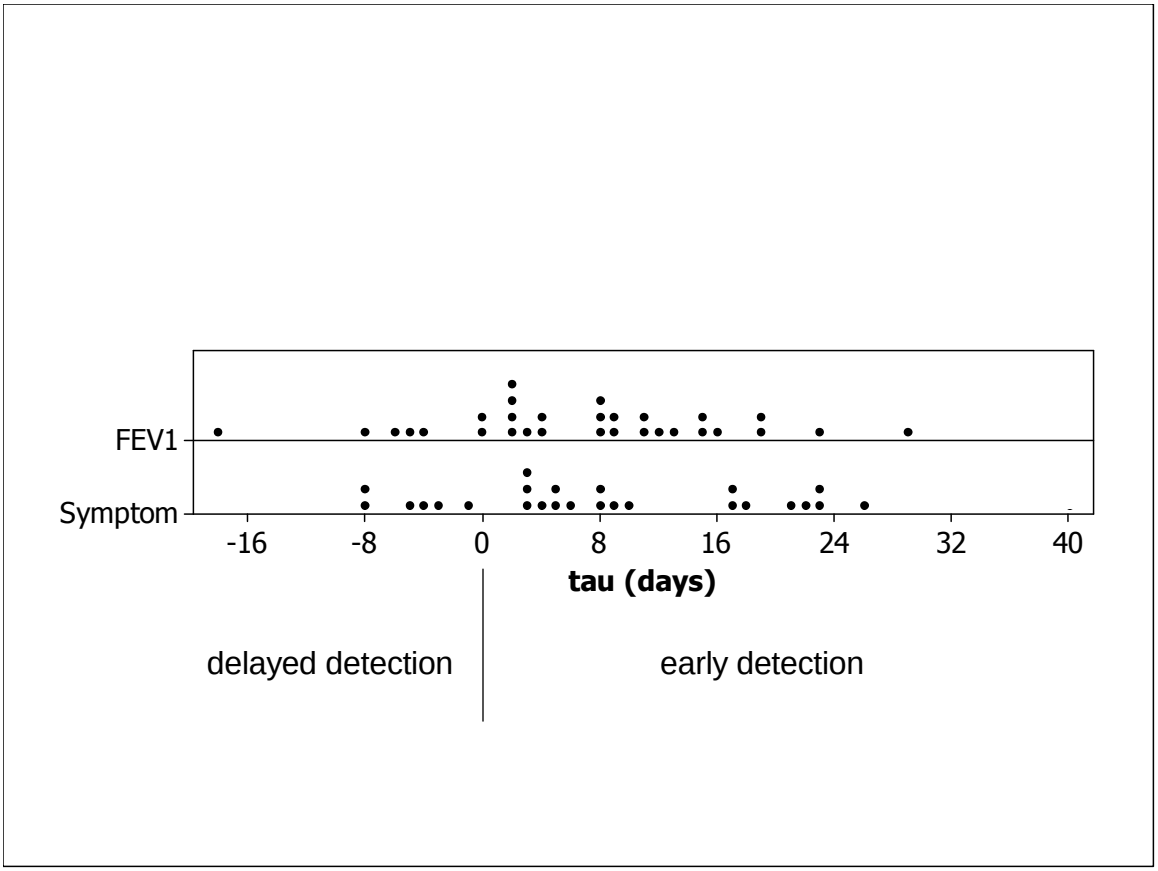

Figure 34. The distribution of timing of detection τ (days), validation set.

Summary

Wavelet-based multiresolution signal crossover and its directional one-sided CUSUM is feasible to detect spirometry decline or symptom increase associated with acute pulmonary events. These methods are data driven and can execute automatically. Symptom and FEV_1 achieved top tier overall per pAUC in OC metrics or AUC in ROC metrics, in both the learning and validation datasets. The results from the validation set are in general, consistent with results in the learning set, suggesting the detection method, designed based on the learning set, can be implemented on new, previously unseen data.

Chapter V

DEMPSTER-SHAFER CLASSFIFIER COMBINATION FOR EVENT DETECTION

Classifier Combination Approaches

Classifier combination aims to achieve a higher level performance than that of each individual classifier. Combining classifiers may improve efficiency and reliability in classification. This is because the sets of data misclassified by the different individual classifiers would not necessarily overlap. In LTHMP, post transplant rejections or infections can originate from different etiologies such as small bronchial airways, mid airways or large airways and may be manifest with symptoms, thus FEV_1, MEFR, PEFR, FEV_1/FVC ratio and symptoms may potentially offer complementary information, which if combined in an effective way, could improve the performance not possible with each individual classifier. Therefore, this stage extends from a single-input-based event classification into a combination-based event classification.

Classifier combination has been an active topic in the pattern recognition community. Majority voting, neural network, Bayesian and the Dempster-Shafer schemes have been proposed with demonstrated successes [143, 144, 145, 146,147]. The common fixed-rule approaches include majority voting for binary label outputs, and maximum, minimum, median and average/linear combination for continuous measurement outputs.

In particular, Bayesian formalism is a probabilistic model-based approach which is now canonical in probabilistic modeling. However, there are situations where the classical Bayesian method is difficult to apply such as lack of proven probability distribution model or unavailability of accurate mathematical analysis [148]. Because of the challenges in modeling classifier output for time series signals, this research looks into rules based on the Dempster-Shafer theory.

Dempster-Shafer's Theory of Evidence

Dempster-Shafer theory of evidence (DST) is a tool for representing and combining measures of evidence. DST is founded on the work of Dempster (1968) [149] and Shafer (1976) [150]. According to Shafer and Pearl [151], DST "provides a non-Bayesian way of using mathematical probability to quantify subjective judgment. Whereas a Bayesian assesses probabilities directly for the answer to a question of interest, a belief function assesses probabilities for related questions and then considers the implications of these probabilities for the question of interest". DST is a generalization to the Bayesian reasoning with more flexibility when the knowledge is incomplete and one has to deal with uncertainty and ignorance. DST "have been popularized in the literature on Artificial Intelligence (AI) and Expert Systems, with particular emphasis placed on combining evidence from different sources" [152]. This is particularly desirable when information from two different apparatus, spirometry and symptom is used in LTHMP. Many engineering sensor fusion problems resort to DST to reduce high false alarm rate [153]. The output of several classifiers may be combined, yielding decision procedures very robust to sensor failures [154].

A Mathematical Description of DST

Let Θ be a set of mutually exhaustive and exclusive atomic (not sub dividable) hypotheses where each possible state of a system $\theta_1,\ldots, \theta_N \ni \Theta$. Θ is referred as the "frame of discernment". Each hypothesis, θ_i is assigned a degree of belief called the "basic probability assignment" (*bpa*), through a set of belief mass functions, $m(.)$. A mass function $m(.)$ assigns beliefs in a hypothesis H as "the measure of belief that is committed exactly to H (Shafer)". *bpa* is subject to the following constraints:

$m(\emptyset)=0$ for an empty set $\emptyset$, and

$m(H)\geq 0$, $\forall H \subseteq \Theta$, and

$$\sum_{H \subseteq \Theta} m(H) = 1 .$$

Each hypothesis that has $m(.)>0$ defines a focal element and their associated *bpa* define a body of evidence (BOE).

A belief function describes the belief in a hypothesis H.

$$Belief(H) = \sum_{B \subseteq H} m(B)$$

for any belief committed to hypothesis B that is part of H. A plausibility function of H is defined as

$$Plausibility(H) = \sum_{B \cap H \neq \emptyset} m(B)$$

and the rule to combine independent evidence $E1$, $E2$, corresponding to belief committed to hypothesis B and hypothesis C, respectively, into a single belief ("$\oplus$"),

$$m_{12} = m_1 \oplus m_2$$

is

$$m_{12}(H) = \frac{\sum_{B \cap C \subseteq H} m_1(B) m_2(C)}{\sum_{B \cap C \neq \emptyset} m_1(B) m_2(C)} . \qquad (14)$$

Notes on the Independence Assumption

DST assumes independence evidence sources, although few applications actually required independence [144,145,146] and it was not considered a major concern [155]. Here, we briefly discuss and justify this independence assumption.

By definition (Figure 3), PEFR and MEFR are flow rates measured at independent sequential time points; FEV_1, FVC are both volume measures with FEV_1 being a fixed-time volume and FVC the total capacity volume is unrelated to time measures. Therefore, these variables are independent measures of different

aspects of a physical maneuver. Their derivatives are reasonably assumed independent as well.

Symptom is measured by a different apparatus from spirometer, which is the subject him/her- self. Based on these considerations, the independence assumption for the Dempster-Shafer rule is reasonably satisfied.

Design of A Dempster-Shafer Classifier

At this point, a single input vector $x \in R^1$ has been processed into a CUSUM vector $v \in R^1$, where $v[t_i] = f(x[t_i])$, at time t_i, i=1,2,3,...,n, and assume v_j are soft labels (as opposed to "hard" labels such as binary classification) of the jth segment of t_i as a continuous measurement of support to event. The classifier output v can reach a binary classification by comparing v to a threshold.

To apply Dempster-Shafer theory to event detection, this research followed the frame work for signal detection by Safranek [156], Boston [157] and Jones [158].

The event detection problem was reduced to just two elements: event present (S) and event absent (N) at any time instant. All possible subset of the set {N, S}, or the power set contains four elements: {N}, {S}, and two additional elements, {N or S} denoted as {U}, the union of N and S, i.e., either event present or noise, and the empty set {ø}. This set Θ={N, S, U, ø} is the "frame of discernment".

Let us select an event detector such as a spirometry variable, x, which contributes its observation by assigning its beliefs over Θ as *bpa*, denoted by a mass function m_i. A *bpa* $m(A)$ represents the portion of total belief that is assigned exactly to set A. *bpa* is subject to the following constraints:

$m_x(ø)=0$

$m_x(S) + m_x(N) + m_x(U)=1$

$m_x(U)$ represents ignorance concerning presence or absence of an event signal.

To combine evidence from both x and y that supports a proposition of interest S, the combination of x's observation m_x and y's observation m_y in DST is

$$(m_x \oplus m_y)(S) = \frac{\sum_{B \cap C = S} m_x(B) m_y(C)}{\sum_{B \cap C \neq \phi} m_x(B) m_y(C)} \tag{15}$$

Next, the problem concerns combining multiple output-level classifiers previously derived. A multi-classifier combination involves these major steps:

1) Select a confidence factor conversion/normalization;
2) Assign mass function to obtain belief and ignorance levels;
3) Calculate combined belief per the Dempster-Shafer Rule;
4) Classification.

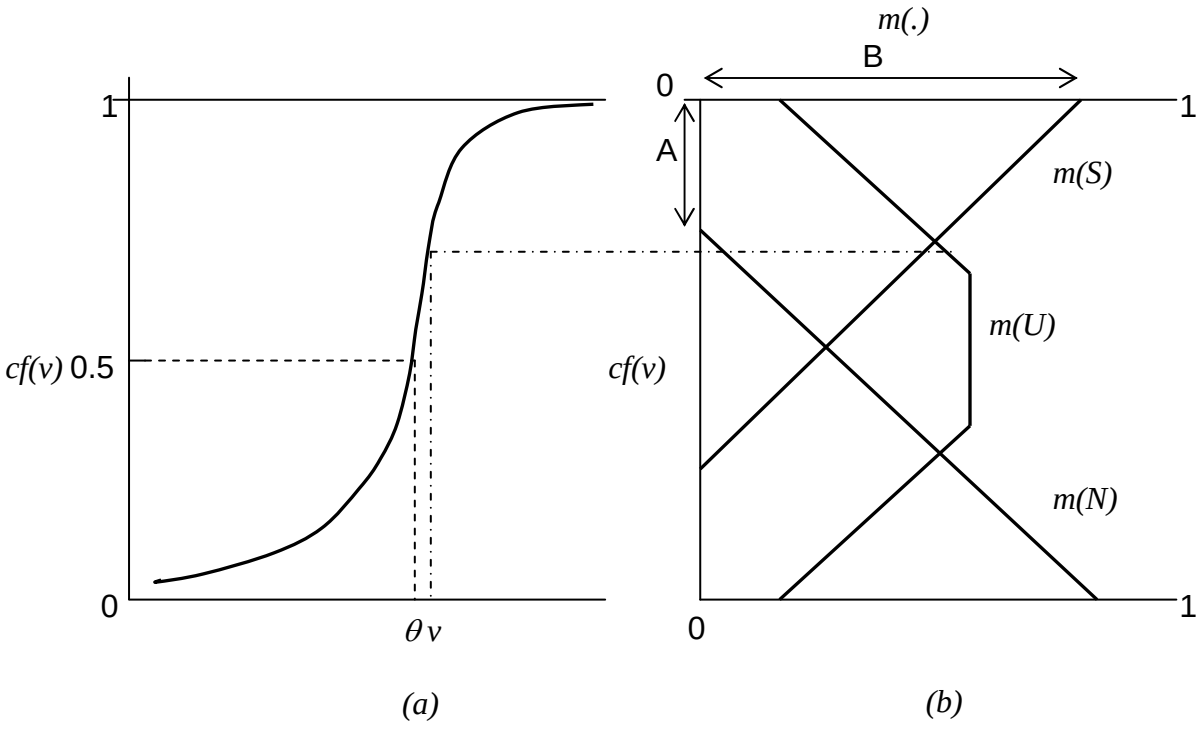

Figure 35. The converter function *cf(v)* and *bpa* assignment procedure.

(a) conversion of input discriminator variable, v, into confidence factor using a sigmoid function; (b) *bpa* mass functions.

Converter Function

The CUSUM, previously derived output in the wavelet approach for a single variable, is used as the input to the combination rules. The first stage of the procedure was to convert the one-sided CUSUM vector v to a confidence factor $cf(v)$, on a scale of 0-1, which represents a level of confidence in (or not in) the variable's support to an event hypothesis. A sigmoid function was proposed by Safranek [156]

$$cf(v) = \frac{1}{1+e^{-k(v-\theta)}} \tag{16}$$

where k describes the steepness of the curve and θ produces a neutral support. Given a hypothesis, S, and its compliment, N, a larger v value implies a larger $cf(v)$ which provides more support to S, while a small v provides more support to N, and vice versa. Let

$$cf(v) = \frac{1}{1+e^{-k(v-\theta)}} \equiv \frac{1}{1+e^{-(a+bv)}} \tag{17}$$

which normalizes an input vector v to $cf(v) \in [0,1]$ and control parameters (k,θ) *or* (a,b). At a threshold cutoff $\boldsymbol{T}$, the converted cutoff value z is

$$z = \frac{1}{1+e^{-(a+b\cdot v)}} \Big|_{v=T}$$
$$z_T = \frac{1}{1+e^{-(a+b\cdot T)}}$$

Continuing from equations (8-9), for the jth CUSUM sequence v_j, $j=1,2,\ldots, J$, select a set of parameters (a,b) such that $OC(v|T) \equiv OC(cf(v)|z,a,b)$. This is to compute a, b such that the mean squared error (MSE) function is minimized, i.e.,

$$(a,b) = arg\ min\ MSE\ \{\ |OC(v_j|T) - OC(cf(v_j)|z,a,b)|\} \tag{18}$$

for all J CUSUM sequences. This was done by iteratively updating (a,b) such that equation (18) convergences.

1. Define a grid of (a_i,b_i), i=1, 2, … I.

2. Select T_k , classify $cf(v_j)$ at a threshold $z_k=cf(T_k)$, k=1,2,…,K.

3. For the jth $cf(v_j)$, compute $\hat{p}_j = max(cf(v_j))$, j=1,2, …, J. (page 65)

$$Y = I(\hat{p}_j \geq z_k)$$
$$\bar{Y} = I(\hat{p}_j < z_k)$$,

Compute x_k=SENS(z_k), y_k=FAR(z_k) for all J per Eq's set (8-9) (page 65)

Compute MSE { |OC(v|T) – OC($cf(v)|z_k,a,b$)|} as

$$\sum_k (x_{k0} - x_{k1})^2 + (y_{k0} - y_{k1})^2 / k$$

Subscript "$_0$"denotes OC(|T) in CUSUM scale, "$_1$" denotes OC(|z_k,a,b) in $cf(v)$ scale.

4. Update a new set of (a_i,b_i), repeat Step 2-3.
5. Select (a_i,b_i) such that $arg\ min\ MSE$ { |OC(v|T) – OC($cf(v_j)|z,a,b$)|}.

Figure 36. Iterative solution for the converter function $cf(v)$.

One set of solution of (k,θ) for the confidence converter function was computed as follows.

Table 15. A set of parameters for the converter function

|CUSUM| as Input Variables	θ	k
FEV_1	0.402	2.49
FEV_1/FVC ratio	0.387	2.58
PEFR	0.380	2.63
MEFR	0.388	2.58
Symptom	0.403	2.48

bpa (Basic Probability Assignment)

Assume a classifier v to differentiate event class membership. We are concerned about combining output information based on the continuous measurement level. Without losing generality, assume the increase of v corresponds to an increased support for event, and vice versa.

The *bpa* mass function per Safranek is a simple piecewise linear function. Per Figure 35 (b), the belief assignments *bpa* is:

$$\begin{aligned} m(S) &= \{\frac{B}{1-A} cf(v) - \frac{AB}{1-A}\} \cdot I_{m(S)} \\ m(N) &= \{\frac{-B}{1-A} cf(v) + B\} \cdot I_{m(N)} \\ m(U) &= \{1 - m(S) - m(N)\} \cdot I_{m(U)} \end{aligned} \tag{19}$$

where

$$I_{m(S)} = \begin{cases} 1, m(S) > 0 \\ 0, m(S) \le 0 \end{cases},$$

$$I_{m(N)} = \begin{cases} 1, m(N) > 0 \\ 0, m(N) \le 0 \end{cases},.$$

$$I_{m(U)} = \begin{cases} 1, m(U) > 0 \\ 0, m(U) \le 0 \end{cases}.$$

The combination ("⊕") of two independent BOEs $m_x(.)$ and $m_y(.)$ are as follows:

$$\begin{aligned} (m_x \oplus m_y)(S) &= \frac{m_x(S)m_y(S) + m_x(U)m_y(S) + m_y(U)m_x(S)}{1 - (m_x(N)m_y(S) + m_x(S)m_y(N))} \\ (m_x \oplus m_y)(N) &= \frac{m_x(N)m_y(N) + m_x(N)m_y(U) + m_y(N)m_x(U)}{1 - (m_x(S)m_y(N) + m_x(N)m_y(S))} \\ (m_x \oplus m_y)(U) &= 1 - (m_x \oplus m_y)(S) - (m_x \oplus m_y)(N) \end{aligned} \tag{20}$$

In theory, DST combination rule is associative and more classifiers can be combined recursively, $m(s) = (m_1(s) \oplus m_2(s)) \ldots \oplus m_k(s), k \in Z$.

The *bpa* functions are related to the decision maker's belief and trust of the input variable. Given a lower confidence of uncertainty, Q_l, and an upper confidence of uncertainty, Q_u, if $Q_l = Q_u = 0$ indicates no uncertainty assigned by the user. Larger Q indicates higher level of uncertainty. Per Safrankek and Jones, an uncertainty confidence interval (Q) [0.4 - 0.9] implies fairly good amount of ignorance inherited in the input variable for belief assignment. Given Q_l and Q_u, and A, B can be assigned as follows:

$$A = \frac{Q_u - Q_l}{1 + Q_u - 2Q_l} \qquad (21)$$
$$B = 1 - Q_l$$

Table 16. *bpa* parameter table

Q_l	Q_u	A	B
0	0	0	1
0.1	**0.2**	**0.1**	**0.9**
0.2	0.3	0.1111	0.8
0.3	0.4	0.7	0.125
0.4	0.9	0.4545	0.6

Note: Parameters in use bold-faced.

The selection of uncertainty levels is left to user's discretion as it is by definition, subjective. Due to the linearity of $m(.)$, sets of similar (Q_l, Q_u) would produce similar classifications as they only modify the magnitude of *bpa* which is offset by adjusting the threshold cutoff ***T'***. A set of low uncertainty $(Q_l,Q_u)=(0.1,0.2)$ corresponding to $(A,B)=(0.1, 0.9)$ was used in this study.

Illustrations

As illustrated (Figure 37), *bpa* mass function $m(s)$ (belief) resembles the one-sided CUSUM sequences (rectified to be positive). The *bpa's,* show "disbelief" through most part of the monitoring. This is intuitive and consistent with that events occur at discrete times.

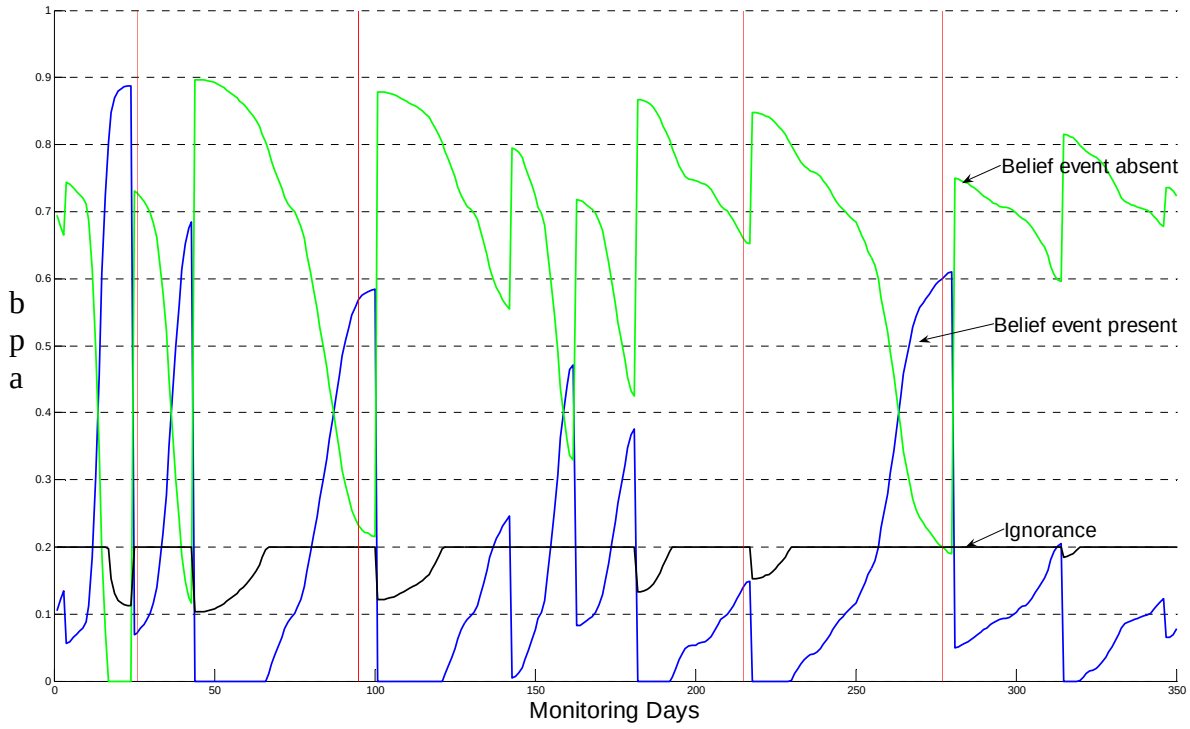

Figure 37. Illustrated *bpa* of FEV_1.

Vertical axis is the *bpa*'s; Horizontal axis is the monitoring days.

The lines represent belief of event ("—"), belief of no event("—"), and ignorance ("—") as labeled in the plot; the vertical lines indicate events ("—").

DS combinations of two individual inputs are first illustrated. The focus is to combine symptom and spirometry.

In Figure 38 (A and B), the input waveforms are the event belief *bpa* series from FEV_1 ($m_x(s)$) and symptom ($m_y(s)$), respectively. Figure 38 illustrates that event labeled ① was detectable by FEV_1 but missed by symptom. One also notices that waveform sequences from the two inputs do not necessarily align in time and their combination is no longer monotonic and may result in truncated waveforms. When both waveforms align, the combination $m_{x,y}(s)$ retains the amplitude of both waveforms. When both the waveforms are small, they do not combine into large

waveforms and are noises to be eliminated. DS combination, per its operation behavior (illustrated in Appendix B), tends to reduce small waveforms and retain larger waveforms. In this case, all events are visually identifiable based on the combined waveforms. In particular, the event misclassified by symptom (labeled ①) is detected by the combination, highlighting the basic benefit when independent detection errors are made by the two classifiers. There are increasing levels of complexity in classification with the sometimes truncated waveforms as a result of combination.

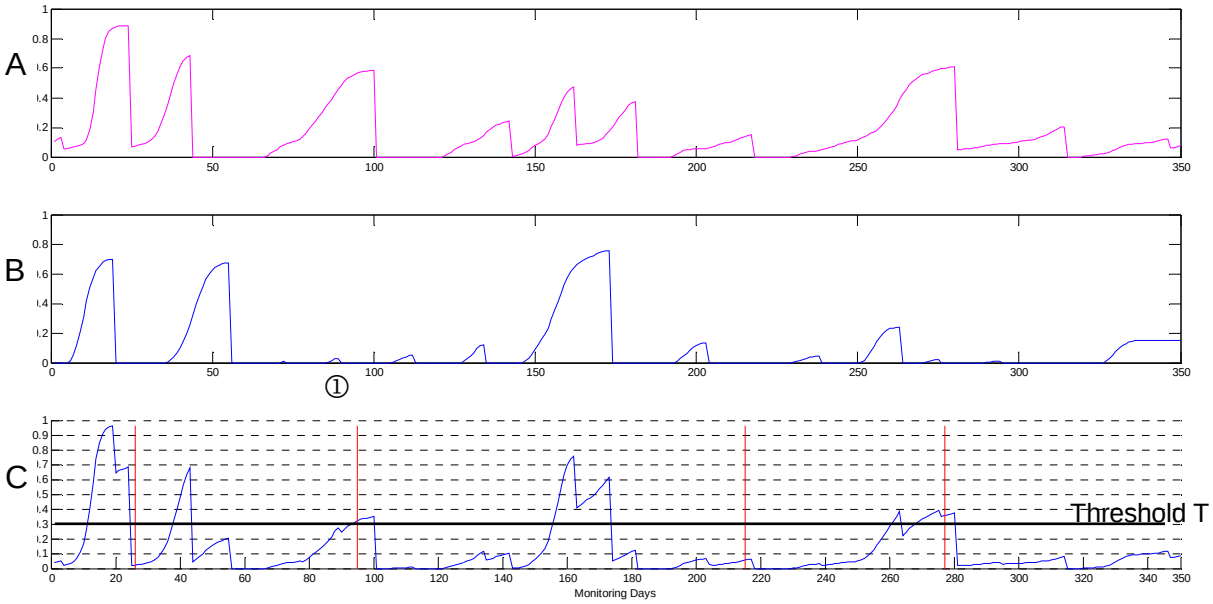

Figure 38. Illustration: DS combination of FEV_1 and symptom.

Horizontal axis is the day of monitoring.

A. *bpa* $m_x(s)$ of FEV_1.

B. *bpa* $m_y(s)$ of symptom.

C. DS combination. Vertical lines indicate events.

(Subject X5183)

In Figure 39, DS rules using two sets of parameters (A,B), denoted as DS1 and DS2, and an average rule are illustrated. DS1 retains the waveforms better than DS2 which is somewhat similar to the average rule. DS1 was used for OC performance evaluations in this research.

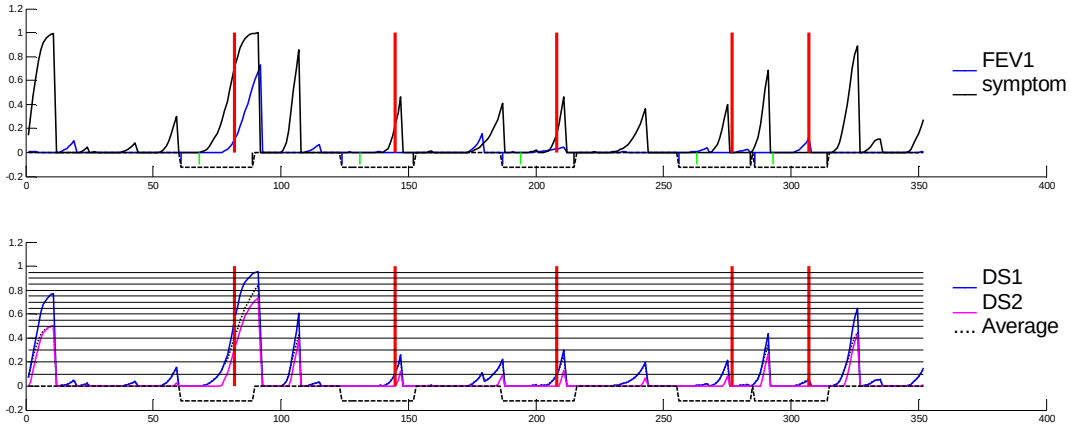

Figure 39. Illustration: Combining FEV_1 and Symptom by DS1 (A=.1, B=.9), DS2 (A=.4545, B=.6), and Average

Horizontal axis is the day of monitoring. Horizontal lines are grid lines.

Vertical lines indicate events.

A 3-week window is plotted as brackets below the horizontal line.

(Subject X5160)

Summary of the Dempster-Shafer Combination

1. Obtain two single classifier time series output vectors $\boldsymbol{v}=v(t_i)$, i=1,2,3,...,n, denoted as $\boldsymbol{x}$ and $\boldsymbol{y}$, respectively, based on previously described procedure in Chapter V.

2. Calculate *bpa* (basic probability assignment)

   a. Normalize $\boldsymbol{v}$ into [0,1] per $cf(v) = \frac{1}{1+e^{-k(v-\theta)}}$ (Figure 36)

   b. Perform *bpa* as

$$m(S) = \{\frac{B}{1-A} cf(v) - \frac{AB}{1-A}\} \cdot I_{m(S)}$$
$$m(N) = \{\frac{-B}{1-A} cf(v) + B\} \cdot I_{m(N)} \quad \text{(Figure 35-b and Eq(19))}$$
$$m(U) = \{1 - m(S) - m(N)\} \cdot I_{m(U)}$$

   where $m(S)$=belief of event, $m(N)$= no belief of event, $m(U)$= ignorance, $I_{m(S)}$, $I_{m(N)}$ and $I_{m(U)}$ are indicator functions, and A,B are scalars for $m(.)$

3. Apply Dempster-Shafer rule below to the two *bpa* vectors $m_x(S)$ and $m_y(S)$ where at each time point t_i, the combined belief is

$$(m_x \oplus m_y)(S) = \frac{m_x(S)m_y(S) + m_x(U)m_y(S) + m_y(U)m_x(S)}{1 - (m_x(N)m_y(S) + m_x(S)m_y(N))} \quad \text{(Eq (20))}$$

4. Obtain an output vector $(m_x \oplus m_y)(S(t_i))$, $i=1,2,\ldots,n$ for classification(page 63)

   (MATLAB code in Appendix D)

Figure 40. Summary of the Dempster-Shafer combination for detection

Results

Detection performance per DS combination was evaluated on both the learning set and the validation set as follows.

OC of DS Combination

DS combinations of two single classifier inputs which include at least one of FEV_1 and symptom (single best variables), are depicted in the following figures.

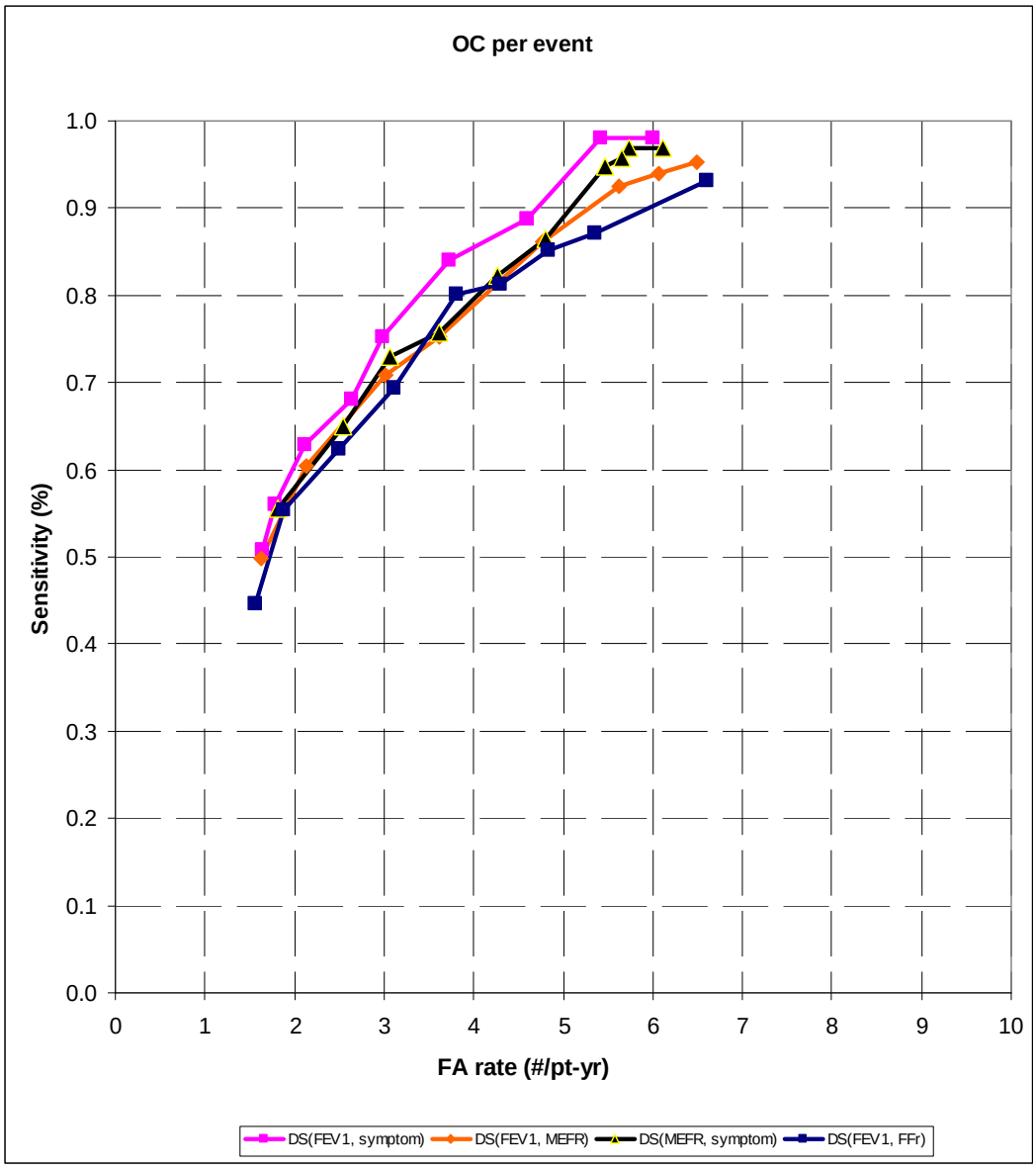

Figure 41. Operating Characteristics of DS combination, per event, learning set. Data table in Appendix C.

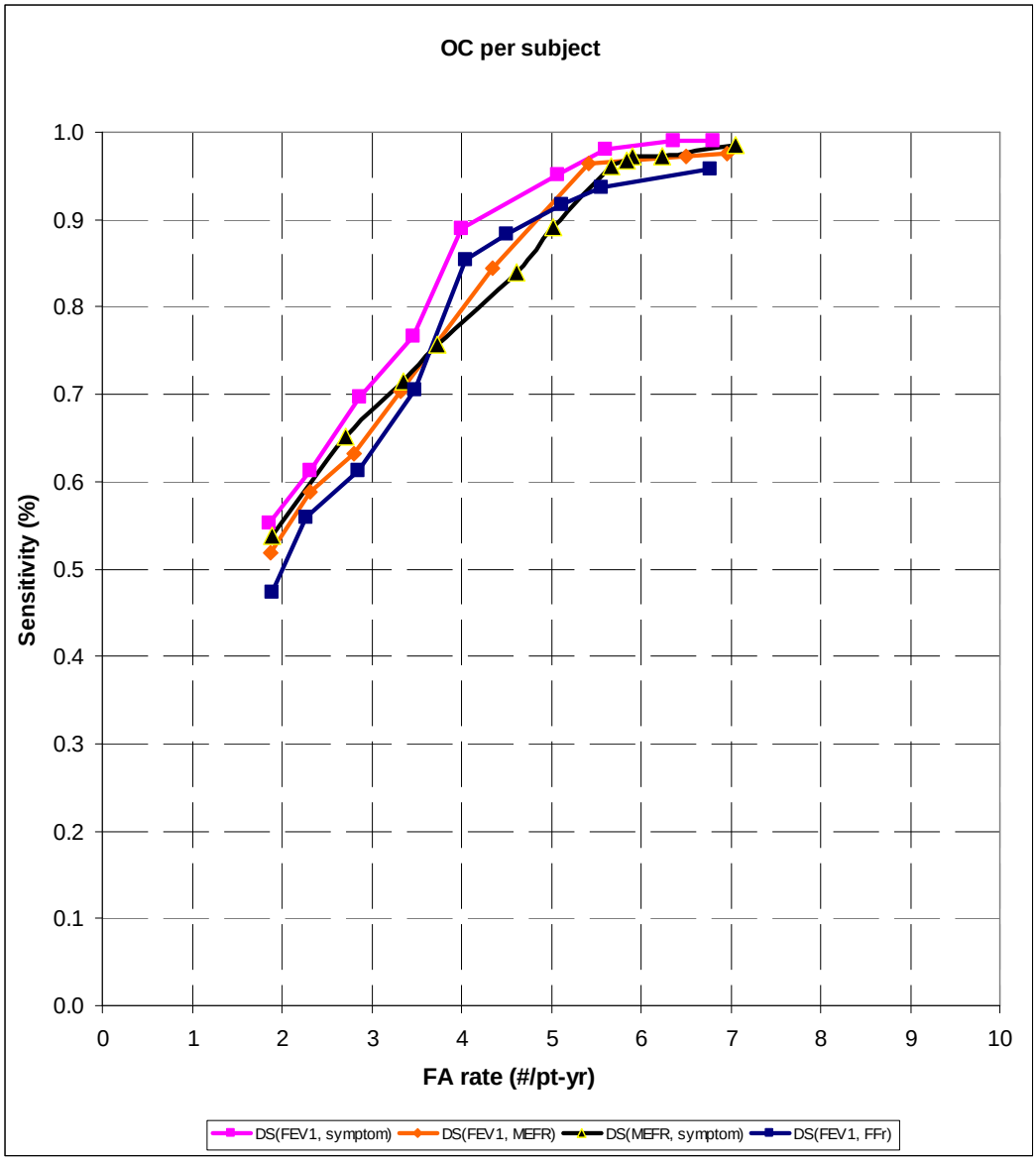

Figure 42. Operating Characteristics of DS combination, per subject, learning set.

Data table in Appendix C.

Similar to the learning set, the OC estimation for the validation set is illustrated below.

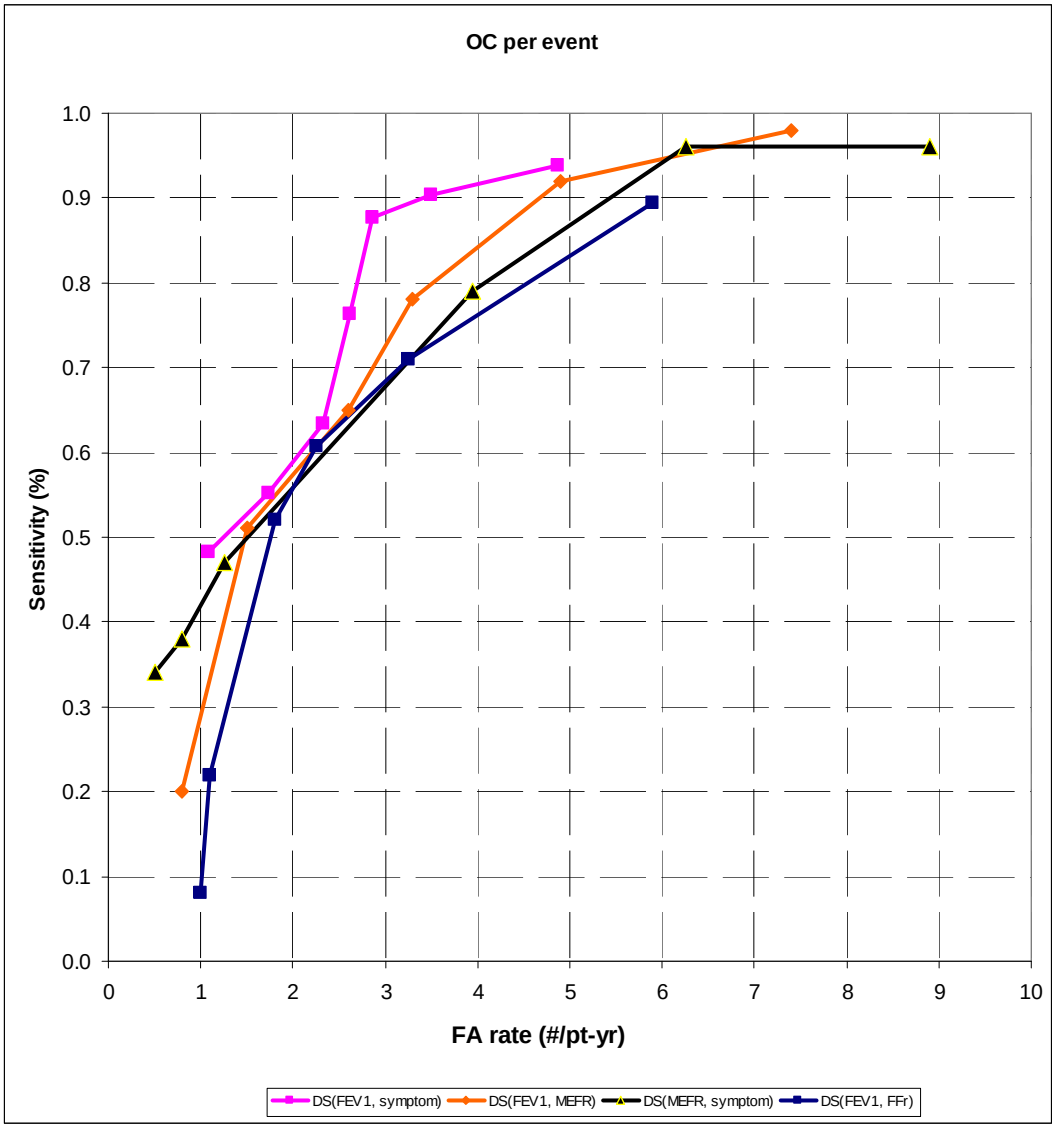

Figure 43. Operating Characteristics of DS combination, per event, validation set.

Data table in Appendix C.

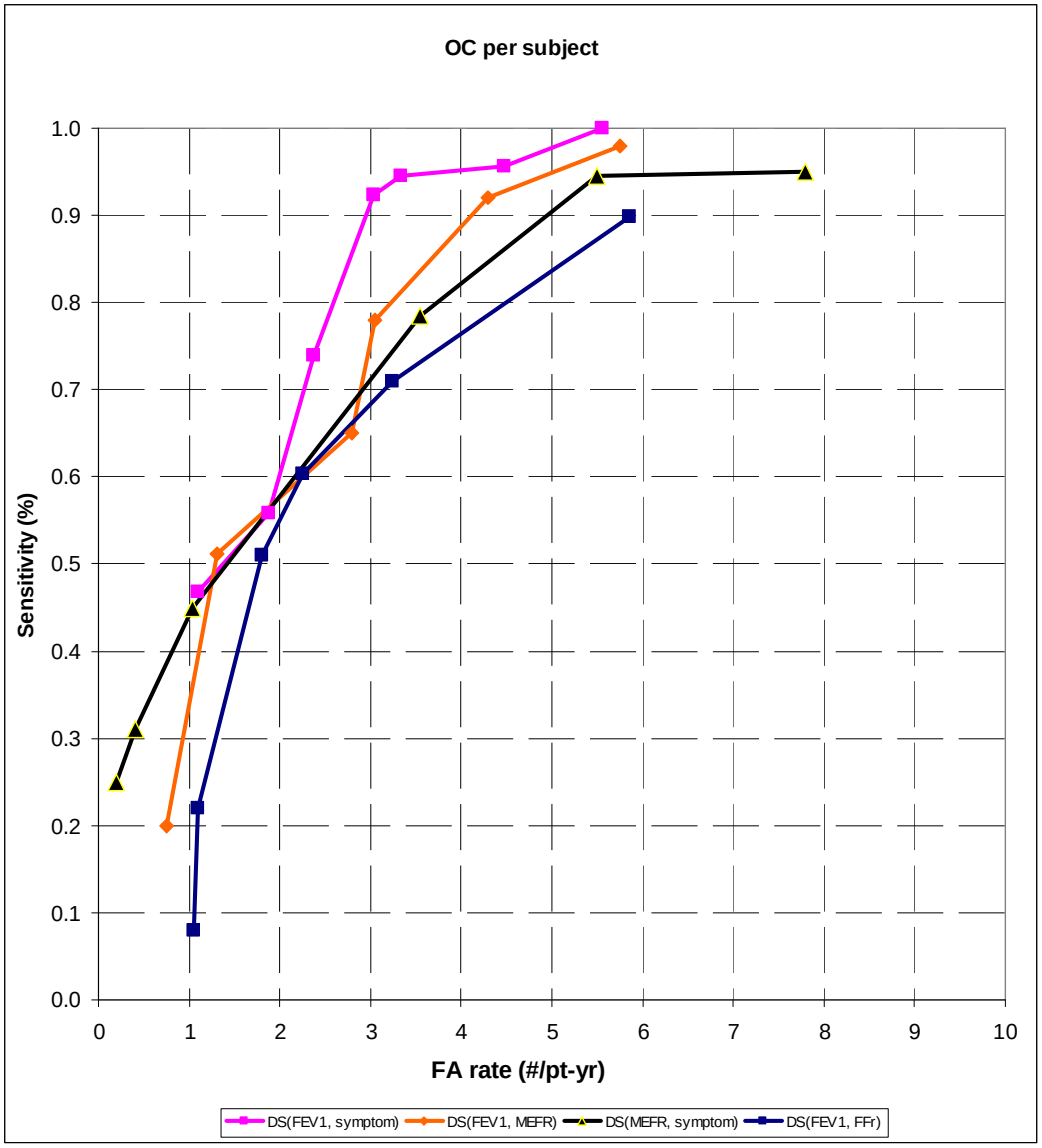

Figure 44. Operating Characteristics of DS combination, per subject, validation set.

Data table in Appendix C.

Per pAUC calculated according to the confining region (Sensitivity: [0.5-1], FAR: [1-6]/subject-yr) and summarized in the table below, two variable Dempster-Shafer (DS) combination candidates are compared.

Table 17. pAUC of individual and DS Combinations

		DS Classifier's pAUC*			
Data Set		FEV_1 & Symptom	FEV_1 & MEFR	MEFR & Symptom	FEV_1 & FEV_1 /FVC
The Learning Set	Per Event	0.545	0.462	0.471	0.438
	Per Subject	0.524	0.463	0.444	0.442
The Validation Set	Per Event	0.634	0.439	0.497	0.358
	Per Subject	0.628	0.503	0.450	0.410

*pAUC – Proportion of partial Area Under the Curve which confines region: Sensitivity [0.5-1] and False Alarm Rate [1-6]/subject-year

Not surprisingly, combination of the two best single classifiers, FEV_1 and symptom, produced the best results in OC and in pAUC. These results also suggest combination involving weaker spirometry variables brings no advantage to further performance improvement.

Results Compared to the Single Input Classifiers

Next, OC curves from the single classifiers of FEV_1 and symptom and their DS combination are compared. Performance metrics previously cited are highlighted with a square box.

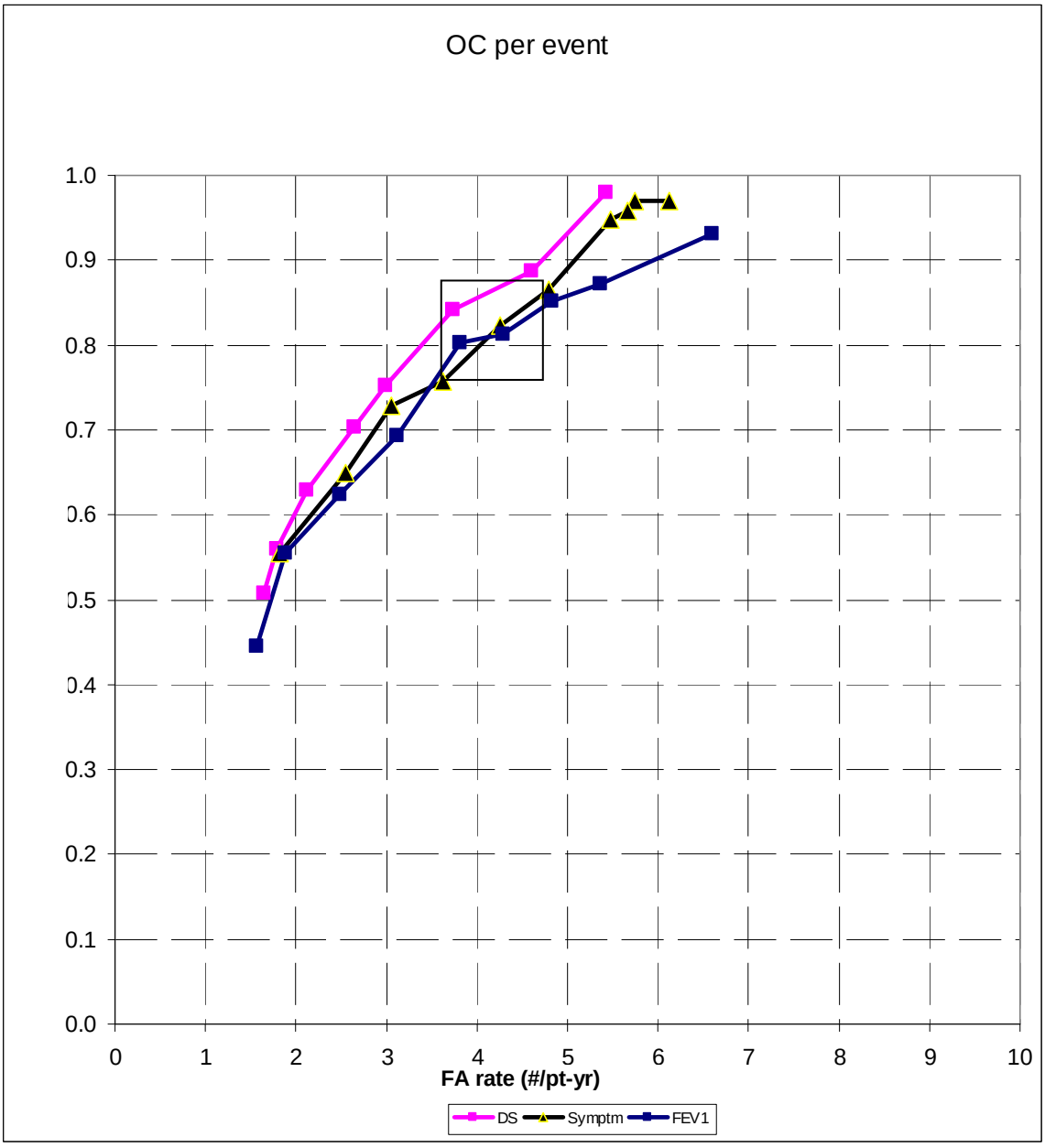

Figure 45. Comparing DS combination vs. single variables, learning set.

Data table in Appendix C.

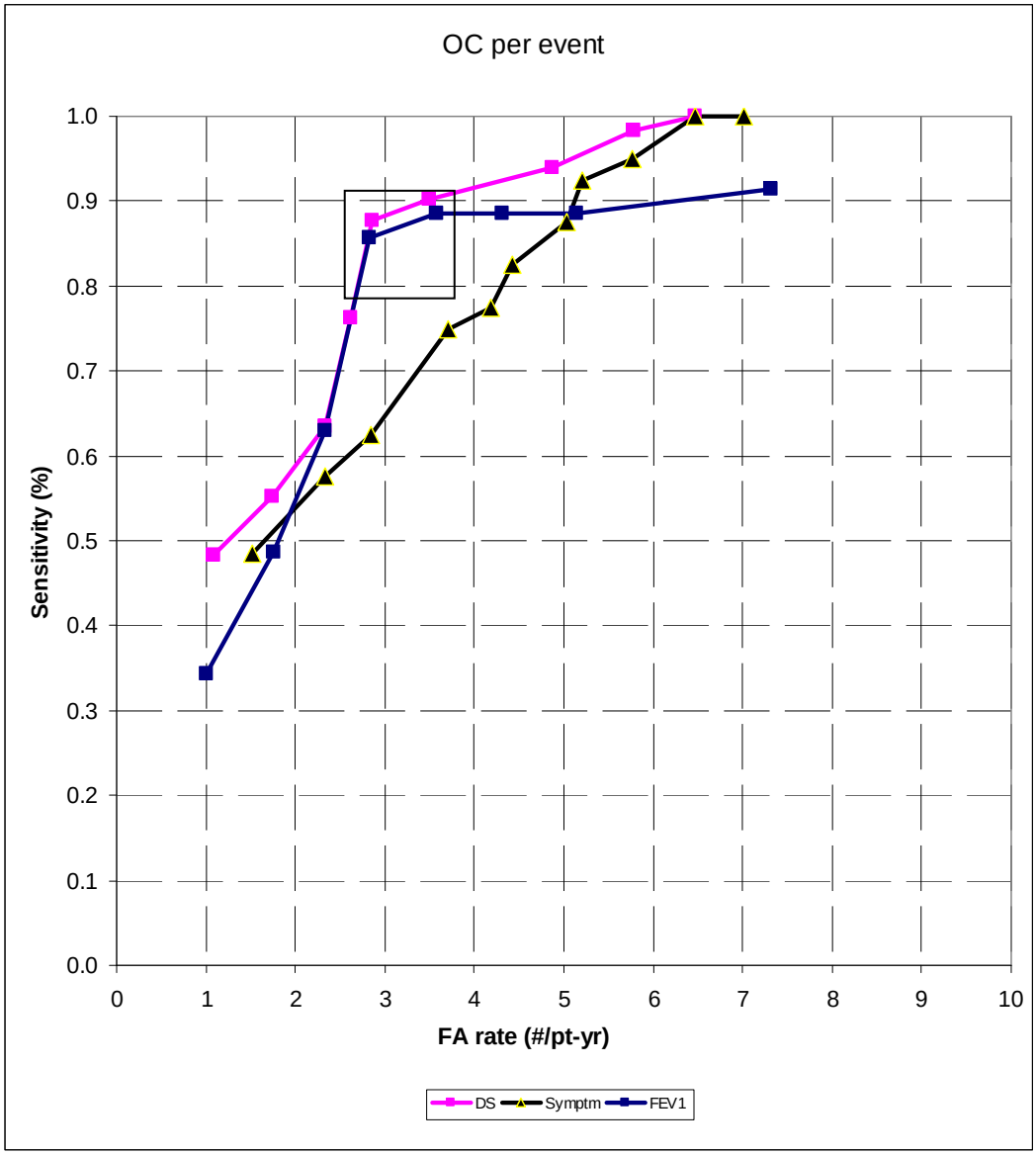

Figure 46. Comparing DS combination vs. single variables, validation set.

Data table in Appendix C.

The OC curve of DS combination is visually above that of the single classifiers. The OC improves by covering more pAUC. For example, OC of (Sensitivity, FAR [/subject-year]) per event were: FEV_1(81.2%, 4.29), Symptom(82.3%, 4.26), and DS(84.1%, 3.73) in the learning set; and FEV_1(85.7%, 2.83), Symptom(87.5%, 5.03), and DS(87.7%, 2.86) in the validation set. The false alarm rates reduced at similar sensitivity. From Table 18, pAUC of DS improved over FEV_1 ranging from 0 - 25%, and improved over symptom ranging from 14 - 49%.

Table 18. pAUC of individual variables and DS Combinations

Data Set		FEV_1	Symptom	DS (FEV_1 & Symptom)
The Learning Set	Per Event	0.433	0.471	0.545
	Per Subject	0.434	0.448	0.524
The Validation Set	Per Event	0.547	0.424	0.634
	Per Subject	0.629	0.549	0.628

*Partial Area Under the Curve region:
Sensitivity [0.5-1] and False Alarm Rate [1-6]/subject-year

Based on these results, DS resulted in an incremental improvement of pAUC over the best individual classifiers.

Summary

The combination of classifiers reduces misclassification errors by individual classifiers from two different apparatus of spirometry and symptom. Dempster-Shafer (DS) combination of FEV_1 and symptom, two best single variable classifiers, provides higher degree of performance than either variable alone, namely, there is increased sensitivity at the same false alarm rate. Per OC metrics, the results are generally consistent between the learning set and the validation set.

Chapter VI

DISCUSSION

In the following, the clinical relevance of this research is discussed first as it is the far most important followed by a review of methodology and factors affecting performance of the algorithm.

Clinical Relevance

Compared to the previous automated triage [17] which assessed the potential of a clinical problem, this research steps further and has evaluated the detection of actual clinical events, based on the same home monitored spirometry and symptoms data. This was realized by customizing methods that fully utilize information embedded in the data, extracting pertinent features, and being guided by knowledge of the occurrence of true clinical events.

The CUSUM classifier performs event detection sequentially. This sequential detection is a significant departure from a fixed sample trial-based detection, which is subject to inflation of false positives with frequent testings (denoted as “multiple testing or multiplicity” problem [159]).

The results show that the best performance was achieved with FEV_1 and symptoms in both datasets. For example of FEV_1, in OC per subject, sensitivity was 85.4% at 4.05 false alarms per subject-year in the learning set and 92.5% at 3.01 false alarms per subject-year in the validation set. The corresponding threshold in OC or ROC point would be set at 0.5 [liter-day] for FEV_1. Sensitivity would increase with higher rate of false alarms at lower thresholds: for example, sensitivity would be 91.8% at 5.56 false alarms in the learning set and 93.8% at 5.70 false alarms per subject-year in the validation set. The thresholds can also be adjusted at user’s discretion to balance preferred sensitivity and tolerance of false

alarms. In addition, these detections were associated with early warning times comparing to the corresponding event records.

In this study, both the learning set and the validation set are from clinical setting, and event detection performance was evaluated on the entire data series of each individual subject which gives a complete assessment of detection performance. Therefore, the event detection results are clinically relevant.

These performance results exceeded reports from clinician-supervised post lung transplant monitoring studies by Morlion *et al* who reported a sensitivity of 63% [18] and Wagner *et al* who reported a sensitivity of 80% with low (but unspecified) specificity [14] both of which were based on daily surveillance of FEV_1 and $FEF_{25-75\%}$(MEFR). In related fields such as colon cancer screening studies, it is also generally considered clinically meaningful to have a sensitivity at or above this range (60% plus) [160]. Therefore, the event detection results are clinically meaningful.

The best overall detection performances were found through FEV_1 and symptoms, however, this interpretation should be confined within the context of acute bronchopulmonary event detection, the targeted event type in this research. It does not deny the importance of other variables, such as the mid flow rate (MEFR), which is considered an early indicator to chronic rejection (i.e., Bronchiolitis Obliterans) [161] and is not a targeted event type in this research.

The use of home monitoring and detection may not preclude a visit to clinic for formal examination and diagnoses. However, more than 80% events would be detected at potentially sooner time. It is also worth noting that healthcare and patient management as sophisticated and costly as they are today, such as detection performance can be achieved by a potentially low cost noninvasive computerized analysis of home spirometry /symptom alone.

It is well known that spirometry alone does not recognize the etiology of pulmonary dysfunction [2,162]. A back pain or a newly diagnosed cancer and treatment observed in our cases may fully explain the worsening of patient's

spirometry and symptom. But distinction between non-pulmonary events and the pulmonary events is not always possible due to the available details in the event annotation. Therefore, this study accounted those otherwise clinically important incidences for “false” alarms due to the strict definition of event. In other words, the true clinically false alarm rate would be lower than estimated.

Factors affecting event detection performance

Any event detection research method would be affected and limited in performance if it encounters critically missing data, especially incomplete or incorrect outcome (event) classifications.

Incomplete Monitoring Data

It has been reported non-adherence to home spirometry was connected to occurrence of events and graft survival [163,164]. In the study datasets, some events might have hindered the subjects from performing home test and consequently, these critical data values were not available. There were learned cases that subjects were reluctant to perform home testings when they expected the results to be poor and were discouraged by that. On the other hand, data missing could be due to “excusable” scenarios such as being recently injured from falling and were not physically feasible to perform home testing; or subjects could have been seen recently at the clinic whose conditions were deemed acceptable; or subjects could have been on vacation which the clinical care staff was made aware of. Therefore, if data are not reported regularly, one first need to rule out that subject’s failure of data transmission is not due to event related decline in health.

Incomplete Event Records

Event record omissions can present a significant problem to rule training, since one basic assumption on training is to have complete and non-contradictory data.

Initially, researchers sourced the learning set events from the medical records at FUMC. Based on initial assessments of functional measurements and symptom diaries, researchers closely scrutinized other FUMC records, looking for any indication of an event. It was estimated that as high as 40 percent of all events might not have been ascertained from the medical record alone [95]. They surmised that FUMC record omissions occurred when patients sought medical attention outside of the Transplant Center for their events. Omissions could then occur because the local doctors failed to communicate event details back to the Transplant Center and / or the Nurse Coordinators at the Transplant Center may have written easily missed cursory notes in the charts following event-informative telephone calls.

The potential impact of missing event records was acknowledged at method selection stage. The method used to derive event features does not require explicit output status at every time instant, and is free from assuming an underlying input-output relation model. In the case of a missed event record, most likely one true positive would have been denied and instead counted as a “false positive”. However, this would penalize the overall false alarm rate proportionally to all candidate variables. Thus performance comparison is not likely to be affected by a few missed events. For example, as in the learning set, assume an additional 20 percent of events were not reported then roughly 40 events would not be included in the 101 events. With 80% sensitivity, 32 events would have been counted as false alarms. That would correspond to 0.86 (32/37.3)/subject-year lower false alarm rate.

Chapter VII

CONCLUSION AND FUTURE WORK

Summary of Contributions

The analysis of post lung transplant home monitoring data is an underserved research area according to the paucity of related publications in the recent years. One key remaining challenge, the detection of actual clinical events, has been addressed in this dissertation research. This research overcomes the limitations in the conventional approaches and exploits numerous cross-discipline methodologies. The wavelet multi-resolution signal analysis, crossover CUSUM for sequential detection, and Dempster-Shafer combination of existing classifiers, each represents the first attempt in this domain literature.

The primary contributions of this dissertation research are:

1. This research represents the first study to carefully characterize long term home monitored data series in LTHMP. Data were demonstrated to be nonlinear, nonstationary, dynamically evolving with long range dependence and contain no known signature patterns readily indicative of events.

2. In this research, a wavelet-based approach was used to analyze time dependent scaling behaviors in data, to re-estimate the underlying signal, and to extract event discriminating features. This data-driven approach is free of any underlying modeling and is applicable to individual level dynamical data.

3. By applying a directional cumulative sum (CUSUM) of multiresolution signal crossover, the event detection scheme has become a sequential testing of change. Sequential detection is advantageous since it may allow for earlier

detection than is possible using trial-based testing with fixed window size. The CUSUM has been demonstrated as an accurate and reliable event indicator. Events can be classified at clinically meaningful sensitivity and specificity, with potential gain of early detection.

4. Dempster-Shafer (DS) multiclassifier combination further reduces the data dimensions and achieves higher sensitivity at fixed false alarm rate than with a single classifier. More importantly, DS provides a unified approach to combine results from different sources or different architectures. Therefore, this is a two stage classification approach extending from univariate classifiers to the combination.

5. The event detection methods, devised under a learning set, demonstrated reproducible results in the validation set. Because both the data sets reflect the clinical setting, the performance results are clinically meaningful. It is recommended to use FEV_1 and symptom for event detection, or their combination for enhanced detection. Because the computation load is manageable under the event detection framework, it has the potential to perform computations in real time.

In conclusion, this dissertation research investigated a set of methods for the detection of true clinical events based on data from clinical settings, and demonstrated promising and clinically meaningful results. The event detection methods are ready to put into practical use.

Future Work

Potential Research. The appropriate data sets containing accurate event information are extremely precious to this research while no public data are available. With more collection of representative and quality data, more stable and representative results based on the same methodology detailed in this research can be expected.

As the healthcare is trending towards personalized analysis [165], the alarm cutoff threshold may be set as adaptive and/ or personalized through a learning process. Machine learning methods including Bayesian modeling can potentially be set up and updated according to learning. This could maximize sensitivity/ minimize false alarm rate at the subject level.

For the classifier combination approach, there are many possibilities: The control parameters in the Dempster-Shafer combination may also be set up as learning parameters, and be personalized to optimize event detection. Other competing combination schemes, such as weighted linear sum [166], averaged Bayes Classifier, or other variations of the Dempster-Shafer combination can be attempted. With large datasets, one can leverage simulation oriented approaches such as bagging and boosting algorithms [167,168] which have been reported with successes to combine weak classifiers [169].

Implementation. To make a real impact, the methodologies from this research need to be implemented in LTHMP. The next step is to test the methods in a prospective clinical setting over a period of time. Computation efficiency, ease of use, real time data screening and cleaning, user interface, and so on, become equally important as the methods themselves.

REFERENCE

1 Levine SM. Can Bronchiolitis Obliterans Syndrome Be Diagnosed by Phone from the Comfort of Home? Chest, Vol.116, pp:5-6, 1999

2 Arcasoy SM, Kotloff RM. Medical Progress: Lung Transplantation. N Engl J Med., Vol. 340, pp. 1081-1091, 1999

3 Taylor DO, Edwards LB, Boucek MM, Trulock EP, Aurora P, Christie J, Dobbels F, Rahmel AO, Keck BM, Hertz MI. Registry of the International Society for Heart and Lung Transplantation: twenty-fourth Official Adult Heart Transplant Report-2007. The Journal of Heart and Lung Transplantation, Vol. 26, pp: 782-95, 2007.

4 Christie JD, Edwards LB, Aurora P, Dobbels F, Kirk R, Rahmel AO, Stehlik J, Taylor DO, Kucheryavaya AY, Hertz MI. The Registry of the International Society for Heart and Lung Transplantation: Twenty-sixth Official Adult Lung and Heart-Lung Transplantation Report-2009. The Journal of Heart and Lung Transplantation - Vol. 28, Issue 10, pp: 1031-1049, 2009

5 Lease ED, Zaas DW. Update on Infectious Complications Following Lung Transplantation. Curr. Opin. Pulm Med. Vol. 17(3), pp: 206-9. 2011.

6 Otulana BA, Higenbottam T, Ferrari L, Scott J, Igbuoka G, Wallwork J. The Use of Home Spirometry In Detecting Acute Lung Rejection and Infection Following Heart-Lung Transplantation. Chest, Vol. 97, pp:353-357, 1990.

7 Bjortuft O, Johasen B, Boe J, Foerster A, Holter E, Geiran O. Daily Home Spirometry Faciliate Early Detection of Rejection in Single Lung Transplant Recipient with Emphysema. Eur Respir J. Vol.6(5), pp:705-708, 1993.

8 Kugler C, Fuehner T, Dierich M, DeWall C, Haverich A, Simon A, et al. Effect Of Adherence To Home Spirometry on Bronchiolitis Obliterans and Graft Survival After Lung Transplantation. Transplantation Vol.15;88(1), pp:129-134, 2009.

9 Otulana B, Higenbottam T, Scott J, Clelland C, Hutter J and Wallwork J, Pulmonary Function Monitoring Allows a Diagnosis of Rejection in Heart-Lung Transplant Recipients. Transplantation Proceedings, Vol. 21, pp:2583-2584, 1989.

10 Higenbottam T, Otulana B and Wallwork J, Transplantation of the Lung. European Respiratory Journal, Vol. 3, pp:594-605, 1990.Inter

11 Adam TJ, Finkelstein SM, Parente ST, Hertz MI. Cost Analysis of Home Monitoring in Lung Transplant Recipients. Int J Technol Assess Health Care, Vol.23, pp. 216-222, 2007.

12 Finkelstein SM, Snyder M, Stibbe SE, Hertz M, et al, Staging of Bronchiolitis Obliterans Syndrome Using Home Spirometry, Chest, Vol.116, pp:120-126, 1999.

13 Ewert R, Wensel R, Muller J, Hetzer R. Telemetric System for an Ambulatory Lung Function Analysis in Transplanted Patients. Transpl Proc. Vol. 32, pp: 204-205. 2000.

14 Wagner FM, Weber A, Park JW, Schiemanck S, Tugtekin SM, Gulielmos V, Schuler S. New Telemetric System for Daily Pulmonary Function Surveillance of Lung Transplant Recipients. Ann Thorac Surg. Vol. 68(6), pp: 2033-2038, 1999.

15 Al-Ashka F, Mehra R, Mazzone PJ. Interpreting Pulmonary Function Tests: Recognize the Pattern, and the Diagnosis Will Follow. Cleveland Clinic Journal of Medicine, Vol. 70, No. 10, 2003.

16 Lindgren BR, Finkelstein SM, Prasad B, Dutta P, Killoran T, Scherber J, et al. Determination of Reliability and Validity in Home Monitoring Data of Pulmonary Function Tests Following Lung Transplantation. Res Nurs Health, Vol. 20, pp:539-50. 1997.

17 Finkelstein SM, M Hertz, et al. Decision Support for the Triage of Lung Transplant Recipients on the Basis of Home-Monitoring Spirometry and Symptom Reporting. Heart & Lung: *The Journal of Acute and Critical Care,* Vol. 34 No. 3, pp 201-8, 2005.

18 Morlion B, Knoop C, Paiva M, Estenne M, Internet-based Home Monitoring of Pulmonary Function after Lung Transplantation. Am J Respir Crit Care Med. Mar 1; 165(5) 694-7, 2002 .

19 Goldberger AL, Amaral LAN, Hausdorff JM, Ivanov PC, Peng CK and Stanley HE, Fractal Dynamics in Physiology: Alterations with Disease and Aging. PNAS Vol. 99, suppl. 1, pp: 2466-2472, 2002.

20 Troiani JS and Carlin BP. Comparison of Bayesian, Classical, and Heuristic Approaches in Identifying Acute Disease Events in Lung Transplant Recipients. Statistics in Medicine Vol. 23, Issue 5, pp: 803 – 824. 2003.

21 Piette JD, Weinberger M, McPhee SJ, Mah CA, Kraener FB, Crapo LM. Do Automated Calls With Nurse Follow-Up Improve Self-Care And Glycemic Controlamong Vulnerable Patients Withdiabetes? Am J Med. 108(1):20-7, 2000.

22 DeNeo DL and Ginns LC. Lung Transplantation at the Turn of the Century. Annu. Rev. Med. Vol.52, pp: 185-201. 2001

23 Trulock EP, Edwards LB, Taylor DO, Boucek MM, Mohacsi PJ, Keck BM and Hertz MI. The Registry of the International Society for Heart and Lung Transplantation: Twentieth Official adult lung and heart-lung transplant report--2003. J Heart Lung Transplant. Vol. 22(6), pp:625-35. 2003.

24 Taylor DO, Edwards LB and Mohacsi PJ *et al*. The Registry of the International Society for Heart and Lung Transplantation Twentieth Official Adult Heart Transplant Report—2003, J Heart Lung Transplant, Vol. 22, pp: 616–624, 2003.

25 Transplant Statistics: 2003 Annual Report. 2003 OPTN/SRTR Annual Report: Transplant Data 1993 - 2002. Table 12. www.ustransplant.org, 2003. Retrieved Jan 18, 2010

26 U.S. Department of Health & Human Services website, OPTN/SRTR Annual Report, http://www.ustransplant.org/annual_reports/current/1216_can-cit_lu.htm, retrieved Feb 15, 2011

27 HHS/HRSA/HSB/DOT. 2009 Annual Report of the U.S. Organ Procurement and Transplantation Network and the Scientific Registry of Transplant Recipients: Transplant Data 1994-2009. U.S. Department of Health and Human Services, Health Resources and Services Administration, Healthcare Systems Bureau, Division of Transplantation, Rockville, MD. 2009.

28 http://optn.transplant.hrsa.gov/latestData/rptData.asp. Retrieved Nov 7, 2011.

29 "Number of U.S. Transplants Per Year, 1988–2008". Infoplease. © 2000–2007 Pearson Education, publishing as Infoplease. http://www.infoplease.com/askeds/3-26-99askeds.html. Retrieved Feb.1, 2011

30 DeMarco T, Frost A, Carrity E, Hertz M, et al, Facts about Lung Transplants, American Society of Transplantation (AST), www.a-s-t.org, 2006. Last retrieved Nov. 10, 2011.]

31 National Heart, Lung, and Blood Institute. Facts About Heart and Heart-Lung Transplants. Bethesda, MD: National Institutes of Health, National Heart, Lung, and Blood Institute], NIH Publication No. 97-2990, Revised August 1997.

32 USC Lung Transplant Program web page, University of Southern California, 2004. http://www.usctransplant.org/lung/. Last retrieved Nov. 10, 2011.

33 Kramer MR, Marshall SE,Starnes VA, Gamberg P, Amitai Z, Theodore J. Infectious Complications in Heart-Lung Transplantation: Analysis in 200 Episodes: Arch Intern Med; Vol.153, pp:2010-6, 1993

34 Dilling DF and Glanville AR. Advances in Lung Transplantation: The Year in Review. The Journal of Heart and Lung Transplantation. Vol. 30, No.3, 2011.

35 Lande JD, Patil J, Li N, Berryman TR, King RA, and Hertz MI. Novel Insights into Lung Transplant Rejection by Microarray Analysis. Proc Am Thorac Soc, Vol 4. pp: 44-51, 2007.

36 Hunter JA, Despines, Wallwork J et al. Heart-lung Transplantation: Better Use of Resources. AM J Med; Vol. 85, pp:4-10. 1988.

37 Shumway SJ and Shumway NE, Thoracic Transplantation, Blackwell Sciences, Inc. 1995.

38 Jamieson SW. Recent Developments in Heart and Heart Lung Transplantation. Transplant Proc; Vol. 27, pp:199-203. 1985.

39 Boehler A, Kesten S, Weder W, et al. Bronchiolitis Obliterans After Lung Transplantation. *Chest.* Vol. 114, pp:1411-1426. 1998.

40 Estenne M and Hertz M. Bronchiolitis Oblitrans after Human Lung Transplantation. Am J Respir Crit Care Med. Vol. 166; pp: 440-444, 2002.

41 Bashshur RL *et al*, Editors, Chapter 2, Clinical Applications in Telemedicine/Telehealth, Telemedicine Journal and E-Health, Vol. 8, No. 1, 2002.

42 Bashshur RL. On the Definition and Evaluation of Telemedicine, Telemedicine Journal, vol. 2, No. 1, pp: 19-30, 1995.

43 Bashshur RL, Mandil SH and Shannon GW. Executive Summary, Telemedicine Journal and e-Health, Vol. 8, No. 1, 2002.

44 Keng S. Health Care Informatics. IEEE Transactions on Information Technology in Biomedicine, Vol. 7, No.1, 2003.

45 Finkelstein SM, Speedie SM, Lundgren JM, Demiris G, Ideker M. TeleHomeCare: Virtual Visits from the Patient Home. Home Health Care Management and Practice Vol. 13, pp:219-226, 2001.

46 Demiris G, Speedie S, Finkelstein SM. The Nature of Communication in Virtual Home Care Visits. Proc Amer Med Informatics Assoc. pp:135-138, 2001.

47 Demiris G, Speedie S, Finkelstein S. Considerations for the Design of a Web-based Clinical Monitoring and Educational System for Elderly Patients. J Am Med Informatics Assoc. Vol. 8, pp: 468-472, 2001.

48 Demiris G, Speedie S, Finkelstein SM. Change of Patients Perceptions of telehomecare. Telemedicine J. and e-Health. Vol.7, pp:241-248, 2001.

49 Veen M, Finkelstein SM, Speedie S, Lundgree JM. Patient satisfaction with Telehomcare. Proc 24th Annual International Conf EMBS (2nd Joint EMBS-BMES Conf) Vol. 24, pp: 1845-1846, 2002.

50 Pangarakis S.J., Harrington K., Lindquist R., Peden-McAlpine C., Finkelstein S. Electronic Feedback Messages For Home Spirometry Lung Transplant Recipients. Heart and Lung: Journal of Acute and Critical Care, 37 (4), pp. 299-307. 2008

51 Laxminarayan S and S. Finkelstein et al., Biomedical Information Technology: Medicine and Health Care in the Digital Future. IEEE Trans. Inform. Technol. Biomed. Vol. 1, pp:1-6, 1997.

52 Laxminarayan S, Information Technology in Biomedicine: Maturational Insights, Vol. 6, No.1, 2002.

53 Kern S and Jaron D. Healthcare Technology, Economics, and Policy: An Evolving Balance. Healthcare Economics, IEEE Engineering in Medicine and Biology Magazine, Jan/Feb 2003.

54 Tom HF Broens, Rianne MHA Huis in't Veld, Miriam MR Vollenbroek-Hutten, Hermie J Hermens, Art T van Halteren and LLambert JM Nieuwenhuis, Determinants of Successful Telemedicine Implementations: a Literature Study. Journal of Telemedicine and Telecare. Vol. 13, pp: 303-309. 2007

55 Geddes LA, Respiration, The Biomedical Engineering Handbook: Second Edition, Ed. Bronzino JD, CRC Press LLC, 2000.

56 R. Pellegrino et al. Interpretative Strategies for Lung Function Tests. Eur Respir J, Vol. 26, pp:: 948–968, 2005.

57 Slonim and Hamilton, Respiratory Physiology, 2nd edition, 1971.

58 Chlan L, Snyder M, Finkelstein S, Hertz M, Edin C, Dutta A. Promoting Adherence to an Electronic Home Spirometry Research Program after Lung Transplantation. Appl. Nurs. Res. Vol. 11(1), pp: 36-40, 1998.

59 Gildea TR and McCarthy K. Pulmonary Function Testing. http://www.clevelandclinicmeded.com/medicalpubs/diseasemanagement/pulmonary/pulmonary-function-testing/ Retrievel Nov. 15, 2011.

60 Quanjer PH et al, Become an Expert in Spirometry, 2005. http://www.spirxpert.com/indices7.htm. Retrieved March 10, 2007.

61 Paggiaro PL, Moscato G, Giannini D, Granco AD, Gherson G, Relationship Between Peak Expiratory Flow (PEF) and FEV1. Eur Respir J. Vol. 10, Suppl. 24, pp: 39s-41s, 1997.

62 Morlion B, Verbandt Y, Paiva M, Estenne M, Michils A, Sandron P, Bawin C, Assis-Arantes P, A Telemanagement System for Home Follow-Up of Respiratory Patients, IEEE Engineering in Medicine and Biology. July/August 1999.

63 S Shingo, J Zhang, TF Reiss. Correlation of Airway Obstuction and Patient Reported Endpoints in Clinical Studies. Eur Respir J. Vol. 17, pp:220-224. 2001.

64 Robinson DR, Chaudhary BA, Speir WA Jr. Expiratory Flow Limitation in Large and Small Airways. Arch Intern Med. Vol.144, pp:1457-1460. 1984.

65 Jain P, Kavuru MS, Emerman CL and Ahmad M. Utility of Peak Expiratory Flow Monitoring. Chest .114.3, 1998.

66 Paradis I, Yousem S and Griffith B, Airway Obstruction And Bronchiolitis Obliterans After Lung Transplantation. Clinics in Chest Medicine. Vol. 14, pp: 751-763, 1993.

67 Dueck R. Assessment and Monitoring of Flow Limitation and Other Parameters from Flow/Volume Loops, Journal of clinical monitoring and computing. Vol. 16, pp: 425-432, 2000.

68 Otulana BA, Higenbottam TW, Scott JP, Clelland C, Igboaka G, Wallwork J. Lung Function Associated with Histologically Diagnosed Acute Lung Rejection and Pulmonary Infection in Heart-Lung Transplant Patients. Am Rev Respir Dis. Vol. 142, pp: 329-332. 1990.

69 Van Muylem A, Melot C, Antoine M, Knoop C, Estenne M. Role of Pulmonary Function in the Detection of Allograft Dysfunction after Heart-Lung Transplantation. Thorax. Vol. 52, pp: 643-647, 1997

70 Becker FS, Martinez FJ, Brunsting LA, Deeb GM, Flint A, Lynch JP III.. Limitations of Spirometry in Detecting Rejection after Single-Lung Transplantation. Am J Respir Crit Care Med. Vol. 150, pp: 159-166, 1994

71 American Thoracic Society. Standardization of spirometry: 1994 update. Am J Respir Cirit Care Med. Vol. 152 (3), pp: 1107-1136. 1995.

72 2004 Annual Report of the U.S. Organ Procurement and Transplantation Network and the Scientific Registry of Transplant Recipients: Transplant Data 1994-2003. Department of Health and Human Services, Health Resources and Services Administration, Healthcare Systems Bureau, Division of Transplantation, Rockville, MD; United Network for Organ Sharing, Richmond, VA; University Renal Research and Education Association, Ann Arbor, MI. 2004.

73 Finkelstein SM and Cady RG, Research Technology: Home Telehealth and Remote Monitoring, Chapter 17, Shaping Health Policy Through Nursing Research, Hinshaw AS and Grady PA (Eds). p. 255-260. Springer. 2011.

74 Goldstein NL, Snyder M, Edin C, Lindgren B, Finkelstein SM. Comparison of Two Teaching Strategies: Adherence to a Home Monitoring Program. Clinical Nursing Research. Vol. 5, No. 2, pp: 150 (17). 1996.

75 Finkelstein SM, Lindgren B, Prasad B, Snyder M, Edin C, Wielinski C, Hertz M. Reliability and Validity of Spirometry Measurements in a Paperless Home Monitoring Diary Program for Lung Transplantation. Heart Lung. Vol. 22(6), pp: 523-33. 1993.

76 De Vito Dabbs A, Hoffman LA, Iacono AT, Zullo TG, McCurry KR, Dauber JH. Are Symptom Reports Useful for Differentiating Between Acute Rejection and Pulmonary Infection After Lung Transplantation? Heart & Lung: The Journal of Acute and Critical Care, Vol. 33, Issue 6, pp: 372-380, 2004

77 Pieczkiewicz DS, Finkelstein SM. Evaluating the Decision Accuracy and Speed of Clinical Data Visualizations." J Am Med Inform Assoc. Vol. 17(2), pp: 178-81. 2010

78 Berry DA and Stangl DK, editors. Bayesian Biostatistics. Marcel Dekker, Inc. 1996.

79 Gelman A. Bayesian Data Analysis. Texts in Statistical Science. Chapman & Hall/CRC Texts in Statistical Science. Second Edition 2004.

80 Carlin B and Louis T. Bayesian Methods for Data Analysis. Chapman & Hall/CRC Texts in Statistical Science, Third Edition. 2009.

81 Allen R, Time series methods in the monitoring of intracranial pressure: Problems, Suggestion for a Monitoring Scheme and Review of Appropriate Techniques. J. Biomedical Engineering. Vol. 5, pp. 5-17, 1983.

82 Melek WW, Lu Z, Kapps A and Fraser WD, Comparison of Trend Detection Algorithms in the Analysis of Physiological Time-Series Data. IEEE Transactions on Biomedical Engineering. Vol., 52, No. 4, 2005.

83 Egan JP. Signal Detection Theory and ROC Analysis. New York: Academic Press. 1975.

84 Basseville M., Nikiforv IV., Detection of Abrupt Changes: Theory and Applicaton. Prentice-Hall, Inc. 1993.

85 http://en.wikipedia.org/wiki/Sequential_probability_ratio_test, retrieved Feb 1, 2011.

86 Wald, Abraham. Sequential Tests of Statistical Hypotheses. Annals of Mathematical Statistics 16 (2): 117–186. June, 1945.

87 Wiklund U, Akay M, Morrison S, Niklasson U. Wavelet Decomposition Of Cardiovascular Signals For Baroreceptor Function Tests In Pigs, IEEE Transactions On Biomedical Engineering, Vol. 49, No. 7, 2002.

88 Karlsson *et al*: Time-Frequency Analysis of ME Signals During Dynamic Contractions, IEEE Transactions On Biomedical Engineering. Vol. 47, No. 2, 2000.

89 Sekine M *et al*. Discrimination of Walking Patterns Using Wavelet-Based Fractal Analysis, IEEE Transactions On Neural Systems And Rehabilitation Engineering. Vol. 10, No. 3, 2002.

90 Saeed M and Mark RG. Multiparameter Trend Monitoring and Intelligent Displays Using Wavelet Analysis. Comput Cardiol. Vol. 27, pp:797-800. 2000.

91 Costa M, Moody GB, Henry I, Goldberger AL. PhysioNet: an NIH a Research Resource for Complex Signals. J Electrocardiol. Vol. 36, pp:139-44. 2003

92 Lakina A, Crovella M and Diot C. Diagnosing Network Traffic Anomalies. Proceedings of ACM Conference of the Special Interest Group on Data Communications (SIGCOMM), 2004.

93 Huang L, Nguyen X, Garofalakis M, Jordan M, Joseph A and Taft N. In Network PCA and and Anomaly Detection. Technical Report No. UCE/EECS-2007-10, UC Berkeley, 2007.

94 Finkelstein SM, Snyder M, Edin-Stibbe C, Chlan L, Prasad B, Dutta P, Lindgren B, Wielinski C, Hertz MI. Monitoring Progress After Lung Transplantation from Home-Patient Adherence. J Med Eng Technol. Vol. 20, pp:203-210. 1996.

95 Troiani JS, Finkelstein SM, Hertz MI, Incomplete Event Documentation in the Medical Records of Lung Transplant Recipients. Progress in Transplantation. Vol. 15, No. 2, 2005.

96 Lande JD, Patil J, Li N, Berryman TR, King RA, and Hertz MI, Novel Insights into Lung Transplant Rejection by Microarray Analysis. Proc Am Thorac Soc. Vol. 4. pp: 44-51, 2007.

97 Box, G. E. P., and Jenkins, G. Time Series Analysis: Forecasting and Control, Holden-Day. 1976.

98 Mallat S, A theory for Multiresolution Signal Decomposition: The Wavelet Representation. IEEE Trans. Pattern Analysis Machine Intelligence. Vol. 11, pp. 674-693, 1989.

99 Meyer Y, Wavelets and Operators. Cambridge, UK: Cambridge University Press, 1992.

100 Daubechies I, Ten Lectures on Wavelets, SIAM-CBMS-NSF Regional Conference Series in Applied Mathematics, 1992.

101 Daubechies I, The Wavelet Transform, Time-Frequency Localization and Signal Analysis. IEEE Trasaction Information Theory. Vol. 36, pp: 961-1005, 1990.

102 Vaidyanathan PP, Multirate Systems and Filter Banks. Englewood Cliffs, NJ: Prentice Hall, 1992.

103 Valens C. A Really Friendly Guide to Wavelets, 1999.

104 Mackenzie D, Daubechies I *et al*. Wavelets, Seeing the Forest – and the Trees Beyond Discovery Beyond Discovery, The Path From Research To Human Benefit, National Academy of Sciences. December 2001.

105 Lewalle J. Tutorial on Continuous Wavelet Analysis of Experimental Data. Syracuse University, 1995.

106 Tewfik AH and Kim M, Correlation Structure of the Discrete Wavelet Coefficients of Fractional Brownian Motion", IEEE Transactions on Information Theory. Vol. 38, pp: 904-909, 1992.

107 Donoho D, De-Noising by Soft-Thresholding. IEEE Transaction on Information Theory. Vol. 41, No. 3, 1995.

108 Anestis Antoiadis, Wavelet Methods in Statistics: Some Recent Developments and Their Applications. Statistical Surveys. Vol. 1, pp:16-55. 2007.

109 Daubechies I, Ten Lectures on Wavelets, SIAM-CBMS-NSF Regional Conference Series in Applied Mathematics, 1992.

110 Goswami JC, Chan AK. Fundamentals of Wavelets. John Willey & Sons, Inc. 1999.

111 Thakor NV, Gramatikov B, Sherman D. Wavelet (Time-Scale) Analysis in Biomedical Signal Processing. The Biomedical Engineering Handbook: Second Edition. CRC Press LLC, 2000.

112 Tao L, Qi L, Zhu S, Ogihara M. A Survey on Wavelet Applications in Data Mining, SIGKDD Explorations. Vol. 4, Issue 2, 2003.

113 Torrence C and Compo GP. A Practical Guide to Wavelet Analysis, Bulletin of the American Meteorological Society, 1998.

114 Mallat S, A Wavelet Tour of Signal Processing, Academic Press, 1997.

115 Staszewski WJ. Intelligent Signal Processing For Damage Detection In Composite Materials. Composites Science and Technology. Elsevier Science Ltd., Vol. 62, pp: 941–950, 2002.

116 Shao X, Model Selection Using Statistical Learning Theory, PhD Thesis, University of Minnesota, 1999.

117 source: http://wavelets.pybytes.com/wavelet/db3/ retrieved Feb 1, 2011.

118 Donoho D, Johstone IM, Ideal Spatial Adaptation by Wavelet Shrinkage, Department of Statistics, Stanford Univeristy, USA, 1993.

119 D. Donoho, De-noising by soft-thresholding. IEEE Transaction on Information Theory, vol. 41, pp. 613-626, May 1995.

120 Johnstone IM and Silverman BW. Wavelet Threshold Estimators For Data With Correlated Noise. Journal of the Royal Sttistical Society. Series B, Vol. 59, No.2, pp: 319-351, 1997.

121 Kovac A and Silverman BW. Extending the Scope of Wavelet Regression Methods by Coefficient-Dependent Thresholding. Journal of the Americal Statistical Association, Vol. 95, No. 449, pp: 172183, 2000.

122 Cai T and Brown LD, Wavelet Shrinkage for Nonequispaced Samples. The Annals of Statistics. Vol. 26, No. 5, pp: 1783-1799, 1998.

123 Ungarala S and Bakshi BR, Multiscale bayesian estimation and data rectification; A. Petrosian, F.G. Meyer, Editors, *Wavelets in Signal And Image Analysis: From Theory To Practice*, *Computational Imaging and Vision*, Springer, Vol. 19, pp. 69-111, 2001.

124 Percival, DB and Walden AT. Wavelet Methods for Time Series Analysis, Cambridge University Press. 2000.

125 Pesquet JC, Krim H and Carfantan H. Time-Invariant Orthonormal Wavelet Representations, IEEE Transactions on Signal Processing. Vol. 44, No. 8, 1996.

126 Wavelet Toolbox, MATLAB 7, The Mathworks Inc., 2005.

127 Donoho D et al, WaveLab and Reproducible Research, Stanford University, 1999.

128 Page ES., Continuous Inspection Schemes. Biomedtrika, Vol. 41, pp: 100-115. 1954.

129 Yang P, Dumont G, and Ansermino JM. A Cusum-Based Multilevel Alerting Method for Physiological Monitoring. IEEE Transactions on Information Technology in Biomedicine, Vol. 14, No. 4, 2010.

130 Fricker RD Jr, Hegler BL and Dunfee DA, Comparing Symdromic Surveillance Detection Methods: EARS' versus a CUSUM-based Methodology. Statistics in Medicine, Vol. 27, pp:3407-3429, 2008.

131 Apiletti D, Baralis E, Bruno G and Cerquitelli T, Read-Time Analysis of Physiological Data to Support Medical Applications, IEEE Transactions on Information Technology in Biomedicine, Vol. 13, No. 3, 2009.

132 Staude GH, Wolf WM, Appel U, and Dengler R, Methods for Onset Detection of Voluntary Motor Responses in Tremor Patients. IEEE Transactions on Biomedical Engineering, Vol. 43, No. 2, 1996.

133 Yu CM, Wang L, et al, Intrathoracic Impedance Monitoring in Patients with Heart Failure: Correlation with Fluid Status and Feasibility of Earl Warning Preceding Hospitalization. Circulation. Vol. 112, pp; 841-848, 2005.

134 Qu H and Gotman J, A Patient-Specific Algorithm for the Detection of Seizure Onset in Long-Term EEG Monitoring: Possible Use as a Warning Device. IEEE Transactions on Biomedical Engineering, Vol. 44, No. 2, 1997.

135 Goense JBM and Ratnam R, Continuous Detection of Weak Sensory Signals In Afferent Spike Trains: The Role Of Anti-Correlated Interspike Intervals In Detection Performance. J Comp Physio A. Vol. 189, pp: 741-759, 2003.

136 Castanho MJP, Barros LC, Yamakami A, and Vendite LL. Fuzzy Receiver Operating Characteristic Curve: An Option to Evaluate Diagnostic Tests. IEEE Transactions on Information Technology in Biomedicine, Vol. 11, No.3, 2007.

137 Winterhalder M, Maiwald T, Voss HU, Aschenbrenner-Scheibe R, Timmer J, Schulze-Bonhage A. The Seizure Prediction Characteristic: A General Framework to Assess and Compare Seizure Prediction Methods. Epilepsy Behav; Vol.4, pp: 318-25. 2003

138 Schelter B, Winterhalder M, Maiwald T, Brandt A, Schad A, Schulze-Bonhage A, et al. Testing Statistical Significance of Multivariate Time Series Analysis Techniques for Epileptic Seizure Prediction. Chaos; Vol.16,pp: 013-108. 2006

139 Goin JE and Haberman JD. Comments on the Logistic Function in ROC Analysis: Applications to Breast Cancer Detection. Meth. Inform. Med. Vol.21, pp: 26-30. 1982.

140 Zhang DD, Zhou XH, Freeman DH and Freeman JL, A Non-Parametric Method for the Comparison of Partial Areas Under ROC Curves and Its Application to Large Health Care Datasets. Statistics in Medicine, Vol.21, pp: 701-715. 2002.

141 Swets JA and Pickett RM. Evaluation of Diagnostic Systems: Methods from Signal Detection Theory. Academic Press, New York, NY. 1982.

142 Egan JP. Signal Detection Theory and ROC analysis. Academic Press, New York, NY. 1975.

143 Kittler J, Hatef M Duin RPW and Matas J. On Combining Classifiers. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 20, No.3, 1998.

144 Kuncheva LI. Combining Pattern Classifiers: Methods And Algorithms. John Wiley & Sons, Inc. 2004

145 Xu L, Krzyzak A, and Suen CY. Methods of Combining Multiple Classifiers and Their Applications to Handwriting Recognition. IEEE Transactions on Systems, Man, and Cybernetics, Vol. 22, No. 3, 1992.

146 Rogova G. Combing the Results of Several Neural Network Classifiers. Neural Networks, Vol.7, No.5, pp. 777-781, 1994.

147 Mandler E and Schumann J. "Combining the Classification Results of Independent Classifiers Based on the Dempster-Shafer Theory of Evidence". Gelsema E and Kanal L (Editors), *Pattern Recognition and Artificial Intelligence*, pp. 381-393. North-Holland. 1988.

148 Abidi MA and Gonzalez RC. Data Fusion in Robotics and Machine Intelligence. Academic Press Inc. 1992.

149 Dempster AP. A Generalization of Bayesian Inference. Journal of the Royal Statistical Society Series B, Vol. 30 (2), pp: 205-247, 1968.

150 Shafer G. A Mathematical Theory of Evidence. Princeton University Press, Princeton, 1976.

151 Shafer G, Pearl J. Readings in Uncertain Reasoning. Kaufmann, San Mateo, CA, 1990.

152 Beynon M, Curry B, Morgan P, The Dempster-Shafer Theory of Evidence: An Alternative Approach to Multicriteria Decision Modeling, Omega, Vol.28, pp: 37-50. 2000.

153 Hutchinson SA and Kak AC. Multisensor Strategies Using Dempster-Shafter Belief Accumulation, in Data Fusion in Robotics and Machine Intelligence, MA Abidi and RC Gonzalez, editors, Academic Press, 1992.

154 Denaeux T, A Neural Network Classifier Based on Dempster-Shafer Theory, IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, Vol. 30, No.2, 2000.

155 Altincay H. On the Independence requirement in Dempster-Shafer Theory for Combining classifiers Providing Statistical Evidence. Appl. Intell., Vol.25, pp: 73-90, 2006.

156 Safranek RJ, Gottschlich S, Kak AC. Evidence Accumulation Using Binary Frames of Discernment for Verification Vision. IEEE Transactions on Robotics and Automation Vol. 6 (4), pp: 405-417, 1990.

157 Boston JR. A Signal Detection System Based on Dempster-Shafer Theory and Comparison to Fuzzy Detection. IEEE Transactions on Systems, Man, and Cybernetics-Part C: Applications and Reviews, Vol. 30, No. 1, 2000.

158 Jones L, Beynon MJ, Holt CA and Roy S, An Application Of The Dempster-Shafer Theory of Evidence to The Classification of Knee Function and Detection of Improvement Due to Total Knee Replacement Surgery. Journal of Biomechancis, 2005.

159 Bland JM, Altman DG. Multiple significance tests: The Bonferroni Method. BMJ; pp: 310:170. 1995

160 S. Syngal, E. Fox, C. Eng, R. Kolodner, and J. Garber. Sensitivity and Specificity Of Clinical Criteria For Hereditary Non-Polyposis Colorectal Cancer Associated Mutations in MSH2 and MLH1. J Med Genet, Vol. 37, pp. 641–645, September 2000.

161 Levine S.M. and Bryan C.L. Bronchiolitis Obliterans In Lung Transplant Recipients: The 'Thorn in the Side' of Lung Transplantation. Chest. 107:894-897. 1995.

162 Becker FS, Martinez FJ, Brunsting LA, Deeb GM, Flint A, Lynch JP III. Limitations of Spirometry In Detecting Rejection After Single-Lung Transplantation. Am J Respir Crit Care Med, 150:159-66. 1994.

163 J. Gottlieb, C. Kugler, T. Fuehner, A.R. Simon, T. Welte, Effect of Non-Adherence to Home Spirometry on Graft Survival and Bronchiolitis Obliterans Syndrome (BOS) after Lung Transplantation. The Journal of Heart and Lung Transplantation, Vol. 27, No. 2S, S69, 2006.

164 Kugler C, Gottlieb J, Dierich M, Haverich A, Strueber M, Welte T, Simon A. Significance of Patient Self-Monitoring For Long-Term Outcomes After Lung Transplantation. Clinical Transplantation. Vol. 24, Issue 5, pp: 709–716, 2010

165 Hudson DL and Cohen ME. Diagnostic Models Based on Personalized Analysis of Trends (PAT). IEEE Transactions on Information Technology in Biomedicine, Vol. 14, No. 4, pp 941-948. 2010.

166 Terrades OR, Valveny E, and Tabbone S. Optimal Classifier Fusion in a Non-Bayesian Probabilistic Framework. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 31, No.9, 2009.

167 Breiman L. Bagging Predictors. Machine Learning, Vol.24, pp: 123-140, 1996.

168 Dietterich T. An Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees: Bagging Boosting and Randomization. Machine Learning, Vol. 40, pp. 139-159, 1999.

169 Ji C and Ma S. Combinations of Weak Classifiers. IEEE Transactions on Neural Networks, Vol. 8, No.1, 1997.

APPENDICES

Appendix A. Definitions used in charactering time series data

Sliding window

A rectangular sliding window is used to summarize the data dynamics over time. An overlapping sliding window of size n intercepts fixed data points from a sample sequence and moves by a block size of n_x data points at a time.

Moving Average

A simple moving average is computed by taking the average of all data points contained in a rectangular sliding window with a refreshing rate of n.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x[i]$$

The Automated Triage Algorithm essentially employs such a moving average estimate. A moving average or “rolling average” is one of a family of similar techniques used to analyze time series data. Moving averages are used to smooth out short-term fluctuations, thus highlighting longer-term trends or cycles. These averages are similar to the low-pass filters used in signal processing. The basic assumption is linearity and Gaussian white noise.

Variance

The variance is

$$variance = \frac{1}{n}\sum_{i=1}^{n} (x[i] - \bar{x})^2$$

A moving variance computes the variance of data series contained in the sliding window involving n observations.

Autocorrelation at lag τ

The autocorrelation is

$$corr(\tau) = \frac{\sum_{i=1}^{N-\tau}(x[i]-\bar{x})(x[i+\tau]-\bar{x})}{\sum_{i=1}^{N}(x[i]-\bar{x})^2}$$

Autocorrelation indicates if a monitored variable is random per inspection of autocorrelation plots, of which the first moment (mean) and second moment (variance) changes throughout the sample sequence.

Signal Noise Ratio (SNR)

The signal noise ratio (SNR) is defined as

$$SNR \equiv \frac{|\,\mathrm{V_i}\,|}{|X_i - \mathrm{V_i}\,|}$$

where X_i is the data vector, V_i is the denoised data representation, and "| |" denotes the norm of the data vector, as in MATLAB.

The signal noise ratio is calculated using a sliding window at each window position. The width of the overlapping sliding window is set at $n=20$ and a block size n_x=2.

Appendix B. Demonstration of Event Detection Using the Triage Decision Support Algorithm

As illustrated on the next page, the automated triage algorithm was emulated to show how it would generate alarms. A simple threshold based algorithm is subject to high rate of false alarms. In this example, there were over a total of 200 alarms over two years.

An alarm is issued if the current weekly average FEV1 declines 6% or greater from the maximum of the previous 3 week MA or 2 month monthly MA. There were 23 and 56 alarms prior to the two events, and 151 alarms generated without ending in an event. Note although the algorithm is sensitive enough to issue alarms prior to the two events, it also produces a large number of false alarms in the later period.

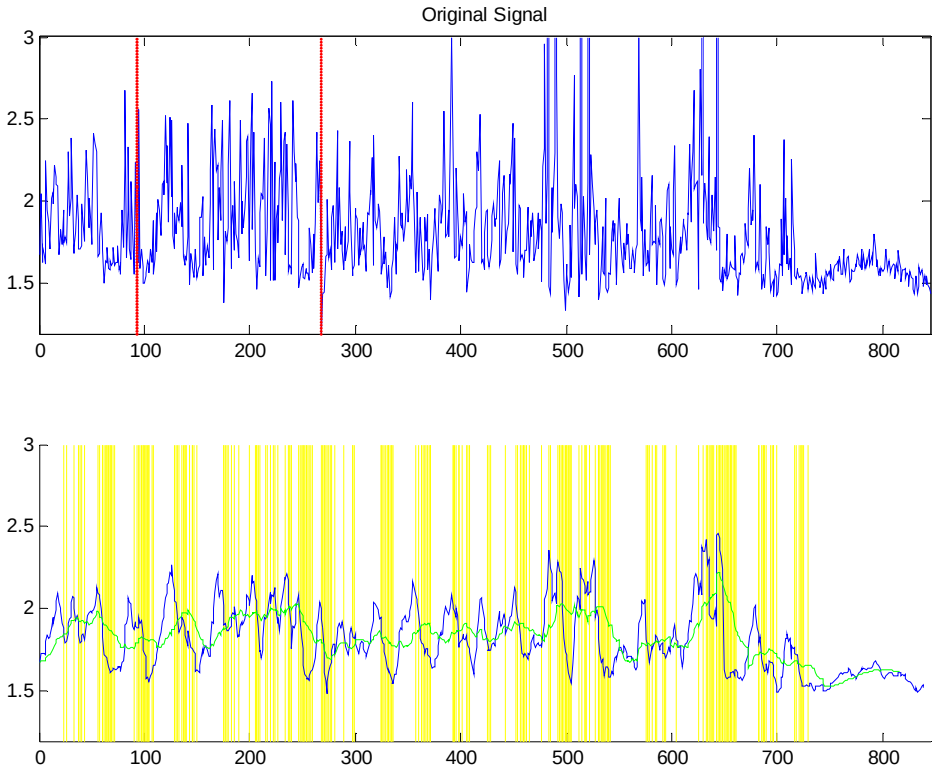

Figure 47. Illustration of a simple threshold-based event detection and alarm generation.

The rule emulates the triage algorithm on daily basis. FEV_1 (Force Expiratory Volume (liters) in one second) is the monitoring variable.

Top: Original FEV_1 signal recordings and event indicator (red lines).

Bottom: Moving Average (MA) (7 day in blue, 30 day in green) and detection alert (yellow lines).

(Subject: X5198)

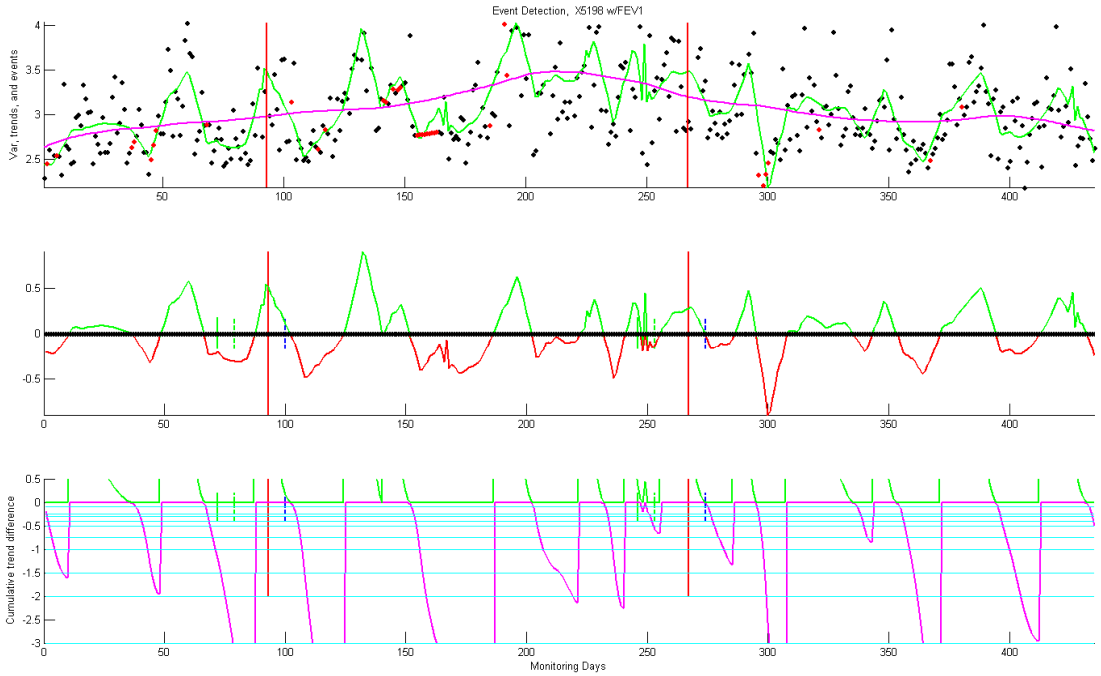

Figure 48. The CUSUM approach derived in this research reduces maximum number of false alarms to 13, of which two sequences detect the events.

Illustration of Simple Combination Rules

For combination at decision level, intuitive simple AND/ OR operation is shown not a good combination rule as illustrated next.

Suppose the inputs are FEV_1 and symptom. As in Figure 49, their CUSUM sequences are plotted and aligned by time. The AND operation would result in positive detection when both sequences exceed their locally set threshold values, shown as the truncation of the CUSUM sequences. The OR operation would be when either of the sequences exceeds its threshold. The resulting combinations are pulse-like sequences. These pulses can be classified per event status as described in Chapter IV-Classification and Performance Evaluation.

From this representative illustration, one learned that the AND operation would reduce event delectability by completely cancelling out inputs if there is disagreement. Indeed, AND operation does not leverage on the compensating potentials of the monitoring variables. The OR operation, on the contrary, will increase event delectability as well as the false alarm rate.

Based on the locally set thresholds $|T|=0.5$ for FEV_1 and $|T|=1$ for symptom, the overall performance evaluation of an OR operation of FEV_1 and symptom yielded a per-event sensitivity of 74% with false alarm rate of 4.6/subject-year, and a per-subject averaged sensitivity of 78% with false alarm rate of 4.9/subject-year.

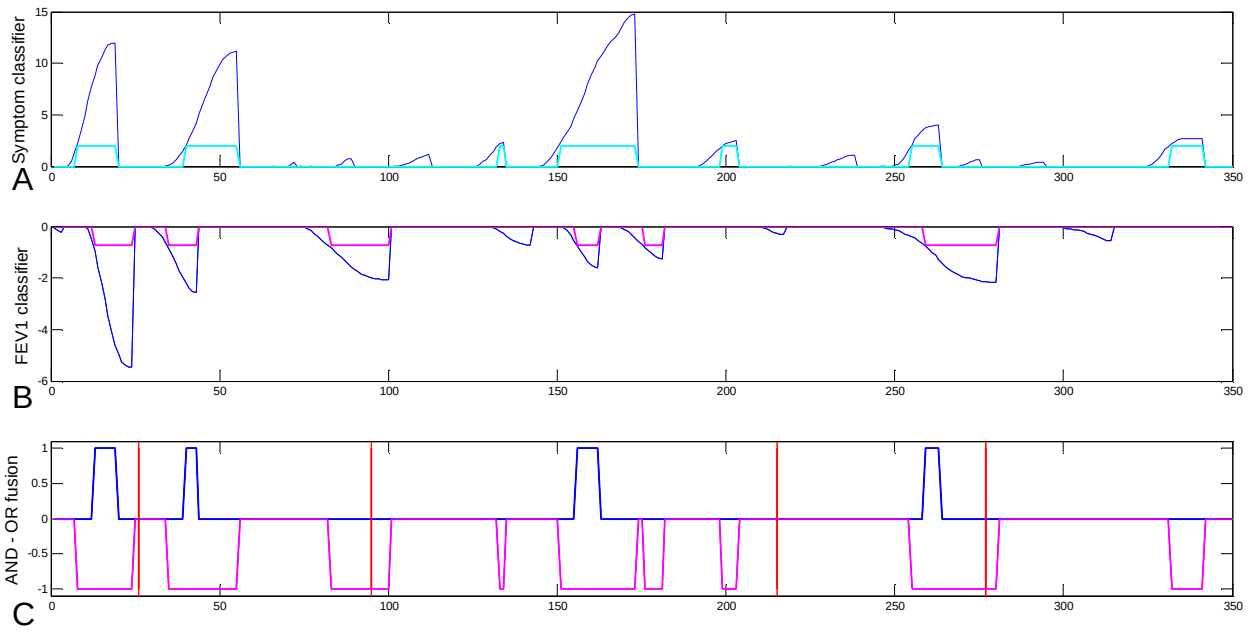

Figure 49. Simple AND/ OR Combination of FEV_1 and symptom

Horizontal axis is the day of monitoring.

A: Symptom CUSUM

B: FEV_1 CUSUM

C: Simple AND (upper) - OR (lower) combination

An Experiment Comparing DS Classifier Combination and Other Fixed Rules

Multiple classifier combination is inspired by the potential of overcoming weakness under a univariate approach. It also has practical utility in the case that critical observations are missing from one variable but available from another variable. Given two input series x and y. The output classifier series are denoted $m_x(s)$ and $m_y(s)$ respectively. Since x and y are independent pieces of evidence, the Dempster-Shafer rule of combination can be applied as:

$$m_{xy}(s) = m_x(s) \oplus m_y(s)$$

Because of the communicative property, the DS combination rule can be iteratively applied to include new classifiers. On classifier combination strategies, other common schemes are majority vote, min, max, median, average (linear sum) and product rules. To illustrate the behavior of a DS rule, two sets of DS rules are compared to the average in Figure 50. The data table below also illustrates the min, max and product rules.

Two input sequences $X, Y \in [0,1]$ are generated to cover the full range of typical values. Two DS rules, i.e., DS1 has no ignorance assigned and DS2 has moderate ignorance assigned. The outputs are plotted as mirror images to the average. Compared to the average, the DS provides flexible combination behaviors. For example, opposing inputs (0.01, 1) gives simple average of 0.5005, vs. DS1 of 1.00 and DS2 of 0.38; high value inputs (0.998, 0.725) gives simple average of 0.86, vs. DS1 of 0.999 and DS2 of 0.748; low value inputs (0.326, 0.16) gives simple average of 0.24 vs. DS1 of 0.086 and DS2 of 0.033; and low value inputs (0.23, 0.11) gives simple average of 0.17 vs. DS1 of 0.036 and DS2 of 0.00. Therefore, the average does not eliminate any inputs. DS2 behaves similar to the average. DS1 operates asymmetrically and cancels out inputs towards zero when both belief values are low; and behaves as an OR/ max operator when at least one of the input values is high. DS has the advantage of generating less ambiguous combination waveforms by cancelation of small values, as opposed to the average

rule. These behaviors are consistent with the design rationale and can reduce false alarm rate at fixed sensitivity.

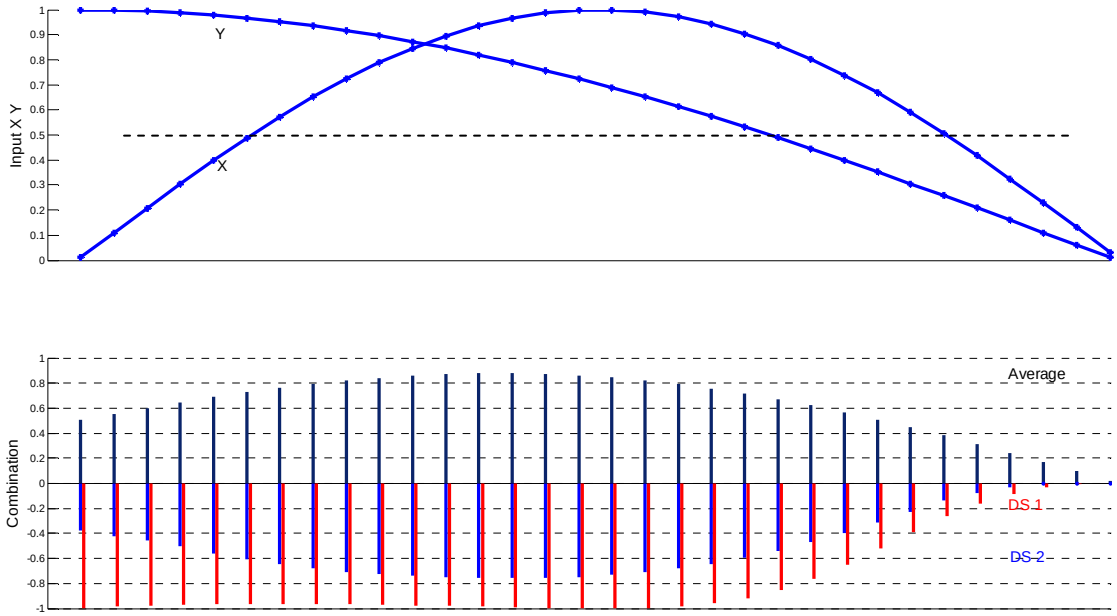

Figure 50. DS combination compared to other rules.

Top: Input sequences X and $Y \in [0,1]$.

Bottom: Average, DS combination (DS 1, A=0.1,B=0.9), and DS combination (DS 2, A=0.25,B=0.6).

Input X	0.010	0.305	0.987	0.998	0.990	0.858	0.590	0.418	0.326	0.230
Input Y	1.000	0.987	0.758	0.725	0.652	0.489	0.306	0.209	0.160	0.111
Minimum	0.010	0.305	0.758	0.725	0.652	0.489	0.306	0.209	0.160	0.111
Average	0.505	0.646	0.873	0.861	0.821	0.673	0.448	0.314	0.243	0.170
Maximum	1.000	0.987	0.987	0.998	0.990	0.858	0.590	0.418	0.326	0.230
Product	0.010	0.301	0.748	0.724	0.645	0.420	0.181	0.087	0.052	0.026
DS 1	1.000	0.971	0.996	0.999	0.995	0.853	0.388	0.160	0.084	0.036
DS 2	0.380	0.504	0.757	0.748	0.710	0.538	0.225	0.081	0.033	0.000

Appendix C. Additional Data Tables

Data tables of Dempster-Shafer combination not listed in main text are included in this section.

Learning set: Figure 41 and Figure 42.

OC performance, per event

	DS (FEV1, symptom)		DS (FEV1, MEFR)		DS (MEFR, symptom)		DS (FEV1, FEV1/FVC)	
$\|T_k\|$	S	FAR	S	FAR	S	FAR	S	FAR
0.025	0.980	6.00	0.952	6.50	0.958	5.66	0.871	5.36
0.05	0.980	5.42	0.940	6.06	0.948	5.47	0.851	4.83
0.1	0.887	4.60	0.925	5.63	0.865	4.80	0.812	4.29
0.15	0.841	3.73	0.861	4.77	0.823	4.26	0.802	3.81
0.2	0.752	2.99	0.752	3.62	0.757	3.62	0.693	3.11
0.3	0.680	2.64	0.709	3.02	0.729	3.06	0.624	2.49
0.4	0.629	2.12	0.605	2.13	0.650	2.55	0.554	1.88
0.45	0.560	1.79	0.498	1.62	0.555	1.82	0.446	1.56
0.5	0.508	1.65	0.952	6.50	0.958	5.6	0.871	5.36
pAUC	0.545		0.462		0.471		0.438	

Note: S: Sensitivity (%event); FAR: False Alarm Rate (#/subject-year).

OC performance, per subject

	DS (FEV1, symptom)		DS (FEV1, MEFR)		DS (MEFR, symptom)		DS (FEV1, FEV1/FVC)	
$\|T_k\|$	S	FAR	S	FAR	S	FAR	S	FAR
0.025	0.990	6.80	0.975	6.96	0.967	5.85	0.936	5.56
0.05	0.990	6.37	0.972	6.50	0.961	5.67	0.917	5.12
0.1	0.980	5.61	0.965	5.42	0.892	5.02	0.884	4.50
0.15	0.951	5.07	0.844	4.34	0.840	4.62	0.854	4.05
0.2	0.890	4.00	0.704	3.33	0.757	3.74	0.705	3.48
0.25	0.766	3.47	0.632	2.81	0.714	3.35	0.612	2.85
0.3	0.697	2.86	0.589	2.32	0.651	2.71	0.559	2.26
0.4	0.613	2.32	0.518	1.88	0.538	1.89	0.473	1.89
0.5	0.553	1.86	0.975	6.96	0.967	5.85	0.936	5.56
pAUC	0.524		0.463		0.444		0.442	

Note: S: Sensitivity (%event); FAR: False Alarm Rate (#/subject-year).

Validation set: Figure 43 and Figure 44.

OC performance, per event

	DS (FEV1, symptom)		DS (FEV1, MEFR)		DS (MEFR, symptom)		DS (FEV1, FEV1/FVC)	
$\|T_k\|$	S	FAR	S	FAR	S	FAR	S	FAR
0.05	1.000	--	1.0000	--	1.000	--	1.000	--
0.1	0.927	4.87	1.0000	--	1.000	--	1.000	--
0.15	0.901	3.50	0.975	7.40	0.950	8.90	0.900	5.90
0.2	0.877	2.86	0.925	4.90	0.950	6.25	0.700	3.25
0.25	0.776	2.62	0.900	3.30	0.800	3.95	0.600	2.25
0.3	0.627	2.33	0.750	2.60	0.450	1.25	0.550	1.80
0.4	0.551	1.74	0.550	1.50	0.400	0.80	0.250	1.10
0.5	0.476	1.09	0.200	0.80	0.350	0.50	0.075	1.00
pAUC	0.634		0.439		0.497		0.358	

Note: S: Sensitivity (%event); FAR: False Alarm Rate (#/subject-year); "-" not evaluated.

OC performance, per subject

	DS (FEV1, symptom)		DS (FEV1, MEFR)		DS (MEFR, symptom)		DS (FEV1, FEV1/FVC)	
$\|T_k\|$	S	FAR	S	FAR	S	FAR	S	FAR
0.05	1.000	--	1.000	--	1.000	--	1.000	--
0.1	1.000	5.56	1.000	--	1.000	--	1.000	--
0.15	0.957	4.48	0.980	5.75	0.950	7.80	0.898	5.85
0.2	0.946	3.33	0.920	4.30	0.945	5.50	0.710	3.25
0.25	0.924	3.04	0.780	3.05	0.785	3.55	0.603	2.25
0.3	0.740	2.37	0.650	2.80	0.450	1.04	0.51	1.80
0.4	0.559	1.87	0.511	1.30	0.310	0.40	0.22	1.10
0.5	0.468	1.09	0.200	0.75	0.250	0.20	0.08	1.05
pAUC	0.628		0.503		0.450		0.410	

Note: S: Sensitivity (%event); FAR: False Alarm Rate (#/subject-year); "-" not evaluated.

Appendix D. MATLAB Codes

```
%function []=MA7baseF(DelThresh, PTmat)
%define threshold: 0.85, 0.93
decline=0.85;

Xpt=X5198;
Xpt=X5198; Xpt(511,:) = X5198(510,:);
ev=[93 267];

%   x0=Xpt(:, 2)./Xpt(:, 3);
eq= [Xpt(:,1)>=Xpt(:,2)];
Xpt(:,1)=Xpt(:,1)-eq.*Xpt(:,1)*0.09;
x0=Xpt(:, 1);

lev=length(ev);

a=size(Xpt);
n = a(1);

xm7=filter(ones(1,7)/7,1,[(ones(1,6)*x0(1))'; x0]);
xm30=filter(ones(1,30)/30,1,[(ones(1,29)*x0(1))'; x0]);
    xm7=[xm7(7:n); xm7(n)*ones(1, 6)'];
    xm30=[xm30(30:n); xm30(n)*ones(1, 29)'];

figure(1);clf
subplot(3, 1, 1);
    signal=x0;
    plot(signal);
    miny=min(signal);
    maxy=max(signal);

    axis([0 n miny maxy]);
    hold
    title('Original Signal');

    for i=1:lev
        plot(ev(i), [miny:(maxy-miny)/100:maxy], 'r-')
        ev(i)
    end
    hold off;


subplot(3, 1, 2);
    hold on;

%run DetectFFr.m to get s
% s0=s(range); s0=s0(2:845);
% plot(s0, 'g-')

%detection loop
```

```
xm7max=zeros(1, n)';
    xm7max(1:7)=max(xm7(1:7));
    for i=8:n
        xm7max(i)=max(xm7(i-6:i));
    end
    xm7=xm7(1:n);
    xm7CMP=[(1:n)' x0 xm7max [x0< (xm7max*decline)]]; %short memory,
compare to 7 days only

    lineX0=xm7CMP(xm7CMP(:, 4)==1);
        disp('lineX0');
    nl=length(lineX0)
    for i=1:nl
        plot(lineX0(i)*ones(1, 2), [miny*1.05; maxy]', 'y-')
    end

%plot data on top of background
    plot(xm7) %remove padding
    axis([0 n miny maxy]);
    hold on;
    plot(xm30, 'g-')
    hold on;

%long term memory
%     xm7=xm7(7:n);
subplot(3, 1, 3);
    hold on;
xm7base=zeros(1, n)';
    xm7base(1:42)=max(xm7(7:14));
    xm7base(43:72)=max(xm7(35:42)) ;
    xm7base(73:180)=max(xm7(65:72));
    xm7base(181:242)=min(xm7(170:180));
    xm7base(243:270)=max(xm7(235:242))  ;
    xm7base(271:287)=min(xm7(260:270));
    xm7base(288:300)=max(xm7(281:287));
    xm7base(301:n)=max(xm7(295:300));

    xm7CMP=[(1:n)' x0 xm7base [x0< (xm7base*decline)]]; %short memory,
compare to 7 days only

    lineX=xm7CMP(xm7CMP(:, 4)==1);
                disp('lineX');
    nl=length(lineX)
    for i=1:nl
        plot(lineX(i)*ones(1, 2), [miny*1.05; maxy]', 'y-')
    end


%plot data on top of background
    plot(xm7) %remove padding
    axis([0 n miny maxy]);
    hold on;
```

```
    plot(xm30, 'g-')
    hold on;

disp('NEW TP FP')
[length(lineX(lineX<93))
length(lineX(lineX> 93 & lineX<267))
length(lineX(lineX>267))]


% Created by Wayne X. Wang
```

Symptom Data Processing

```
if (PT=='3028')
data=data3028Copy;
xi=[2 4 9 10 12:15 17:25 27:36 38:49]'; %tract unfilled indices
end


data1=[data(:, 1) data(:, 10:15)];
% data=data30;
% xi=[]';
Cgh= interp1(data1(:,1),data1(:, 2),xi);
SAmt= interp1(data1(:,1),data1(:, 3),xi);
SCl= interp1(data1(:,1),data1(:, 4),xi);
Whz= interp1(data1(:,1),data1(:, 5),xi);
ShBr= interp1(data1(:,1),data1(:, 6),xi);
Wbg= interp1(data1(:,1),data1(:, 7),xi);

datax=[data1(:, 1:7); [xi Cgh SAmt SCl Whz ShBr Wbg]];
dataxs= sortrows(datax,1);
%mimic baseline in original paper, reset missing value as zero
%Dyspnea
     xShBr=dataxs(:,6); xShBr=xShBr.*[xShBr>0];% x=xb; clf;
TITLE('Dyspnea and Events')
%Wheeze
     xWhz=dataxs(:,5); xWhz=xWhz.*[xWhz>0];% x=xw; clf;TITLE('Wheeze
and Events')

%Sputum (color, amount)
    xSAmt=dataxs(:,3); xSAmt=xSAmt.*[xSAmt>0];
    xSCl=dataxs(:,4); xSCl=xSCl.*[xSCl>0]; %x=xc/3+xa/3;% TITLE('Sputum
(color, amount) and Events')

%Dyspnea+Wheeze+Sputum (color, amount)/3 (weighted sum)
%     x=xb+xw+xc/3+xa/3;clf;TITLE('Dyspnea+Wheeze+Sputum(weighted sum)
and Events')

%Cough
    xCgh=dataxs(:,2); xCgh=xCgh.*[xCgh>0]; %clf; %x=xg;TITLE('Cough and
Events')
```

```
%Relative terms: cough, sputum (amount and color)
    x=(xCgh+xSAmt+xSCl);
%Absolute terms (dyspnea+wheeze)
    x2=xShBr+xWhz; clf;

%freeze current week x data and compare to baseline
%baseline
lxz=length(x);

wi=floor(length(x)/7);

z(1)=0;
if wi>0
z(2)=mean(x(1:1*7));  % for week 2 data
end
if wi>1
    z(3)=mean(x(1:2*7)); % for week 3 data
end
if wi>2
z(4)=mean(x(1:3*7));% for week 4 data
end
if wi>3
z(5)=mean(x(1:4*7));% for week 5 data
end
if wi>4
z(6)=mean(x(1:5*7));% for week 6 data
end
if wi>5
z(7)=mean(x(1:lxz));% for week 7 data
end
if wi>6
z(7)=mean(x(1:6*7));% for week 7 data
end
if wi>7
z(8)=mean(x(1:7*7));% for week 8 data
end

% wi=150; %wi>=9
if wi>8
for i=9:wi
    z(i)=mean(x((i-9)*7+1:(i-1)*7));
end
end

%eg, week i=2 detection
xdiff=x(1:7);

for i=2:wi
    xnow=x((i-1)*7+1:i*7);
    xbase=[ones(7,1)*z(i)];
    xtmp=xnow-xbase;
```

```
xdiff=[xdiff; xtmp];
end
res=(length(x)-7*wi);

if wi>8
    z(wi+1)=mean(x((wi-8)*7+1:wi*7));
    xnow=x(wi*7+1:length(x));
    xbase=[ones(res,1)*z(wi+1)];
    xtmp=xnow-xbase;
end
if wi==7
%     z(wi+1)=mean(x((wi-5)*7+1:wi*7));
    xnow=x(wi*7+1:length(x));
    xbase=[ones(res,1)*z(wi+1)];
    xtmp=xnow-xbase;
end;

xdiff=[xdiff; xtmp];

% absolute term: x=xb+xw+xc/3+xa/3
% relative term: xdiff of cough, sputum (amount and color)
xx=x2+xdiff/2; % absolute + relative scores/2 (weighted)

ssw3028= xx; % save the converted data
```

Signal Noise Ratio using a sliding window

```
%MA estimate
xz=X(lx:-1:1,1);
    xmz7pad=filter(ones(1,7)/7,1, [(ones(1,7)*xz(1))'; xz]); %add 7 day
padding
    xmz7=xmz7pad(8: (lx+7)); %remove padding

    snr0=zeros(1,lx);
    snr=zeros(1,lx);
    eti=zeros(1,lx);
    ns=2;
    for i=2:floor(lx/ns)
        starti=ns*(i-1)+1-ns;
        stopi=ns*i;
        xi=X(1: stopi);
        [V1tmp,CXDV1,LXDV1]   = wden(xi,'sqtwolog','s','one',3,'db3');
            V1(stopi-10+1: stopi)=V1tmp(stopi-10+1:stopi);

        snr0(stopi-10+1: stopi)=norm(xmz7(stopi-10+1:
stopi))./norm(xi(stopi-10+1: stopi) - xmz7(stopi-10+1: stopi));
        snr(stopi-10+1: stopi)=norm(V1(stopi-10+1:
stopi))./norm(xi(stopi-10+1: stopi) - V1(stopi-10+1: stopi));
    end

subplot(211), plot(1:lx, X); hold on; plot(1:lx, V1, 'r'); plot(1:lx,
xmz7, 'm');
subplot(212), plot(1:lx, snr0); hold on; plot(1:lx, snr)
plot(1:lx, snr./snr0)

% Use segments to show statistical significance
ns=20; %non overlapping;
block=floor(lx/ns);
snrb=zeros(1, block);
snrb0=zeros(1, block);
z=zeros(1, block);
   for i=1:floor(lx/ns)
        starti=ns*(i-1)+1;
        stopi=ns*i;

        snrb(i)=norm(V1(starti: stopi))/norm(X(starti: stopi) -
V1(starti: stopi))
         snrb0(i)=norm(xmz7(starti: stopi))/norm(X(starti: stopi) -
xmz7(starti: stopi))
 %Z(i)=(mean(snrb(i))-
mean(snrb0(i)))/sqrt(std(snrb0(i))*std(snrb0(i))+std(snrb(i))*std(snrb(
i)))
   end
   plot(snrb0, snrb, '+')
```

Multiscale analysis

See modwt.m, modwtjm.m in Percival's WMSTA

```
function [Wtout, Vtout] = modwtjm(Vtin, ht, gt, j)
% modwtjm -- Calculate jth level MODWT coefficients (MATLAB
implementation).
%
%****f* wmtsa.dwt/modwtjm
%
% NAME
%   modwtjm -- Calculate jth level MODWT coefficients (MATLAB
implementation).
%
% SYNOPSIS
%   [Wtout, Vtout] = modwtjm(Vtin, ht, gt, j)
%
% INPUTS
%   * Vtin        -- Input series for j-1 level (i.e. MODWT scaling
coefficients)
%   * ht          -- MODWT wavelet filter coefficients.
%   * gt          -- MODWT scaling filter coefficients.
%   * j           -- level (index) of scale.
%
% OUTPUTS
%   * Wtout       -- MODWT wavelet coefficients for jth scale.
%   * Vtout       -- MODWT scaling coefficients for jth scale.
%
% SIDE EFFECTS
%
% DESCRIPTION
%   modwtjm is an implementation in MATLAB code of the MODWT transform
for the jth level.
%
%   the MODWT transform, modwtj, which linked in as a MEX function.
N = length(Vtin);
  L = length(ht);

  Wtout = zeros(N, 1) * NaN;
  Vtout = zeros(N, 1) * NaN;

  for (t = 1:N)
    k = t;
    Wtout(t) = ht(1) * Vtin(k);
    Vtout(t) = gt(1) * Vtin(k);

    for (n = 2:L)
      k = k - 2^(j-1);
      if (k < 1)
        k = k + N;
      end
```

```
      Wtout(t) = Wtout(t) + ht(n) * Vtin(k);
      Vtout(t) = Vtout(t) + gt(n) * Vtin(k);
    end
  end
```

DWT.m, WDEN.m SWT.m see MATLAB7

Compute wavelet regression and CUSUM

```
%WDEN.m is a MATLAB7 function
%  WDEN Automatic 1-D de-noising using wavelets.
%     WDEN performs an automatic de-noising process of a 1-D
%     signal using wavelets.
%
%     [XD,CXD,LXD] = WDEN(X,TPTR,SORH,SCAL,N,'wname') returns
%     a de-noised version XD of input signal X obtained by
%     thresholding the wavelet coefficients.
%     'sqtwolog' for universal threshold sqrt(2*log(.)).
%     SORH ('s' or 'h') is for soft or hard thresholding
%     SCAL defines multiplicative threshold rescaling:
%       'one' for no rescaling.
%       'sln' for rescaling using a single estimation
%         of level noise based on first level coefficients.
%       'mln' for rescaling done using level dependent
%         estimation of level noise.
%     Wavelet decomposition is performed at level N and 'wname'
%     is a string containing the name of desired orthogonal
%     wavelet.

%Getting wavelet estimates V1, V2
        [V1,CXDV1,LXDV1] = wden(X,'sqtwolog','s','one',2,'db3'); %;
        [V2,CXD5V2,LXD5V2] = wden(X,'sqtwolog','s','one',5,'db4');

% Calculating trend difference
V3=V1-V2; %trend difference
V4=V3.*(V3<0); %subzero trend difference
V5=V3.*(V3>=0); %above zero trend difference
V6=V5(lx:-1:1);

% Visualize with events
if (PT=='5183')
    ev=[26 95 215 277];
    lev=length(ev);

    for i=1:lev
        plot((ev(i))*ones(1,2), [minx; maxx], 'r-')
    end
end
% Compute CUSUM (positive and negative)
cum=zeros(lx,1);
cumi=0;
```

```
for i=1:lx
    if V4(i)==0 cumi=0;
        else cumi=cumi+V4(i);
    end
    cum(i)=cumi;
end
plot(cum, 'm-');


cumn=zeros(lx,1);
cumri=0;
for i=1:lx
    if V6(i)==0 cumri=0;
        else cumri=cumri+V6(i);
    end
    cumnr(i)=cumri;
end
cumnn=cumnr(lx:-1:1);
plot(cumnn, '-');
```

Dempster-Shafer Combination

The DS combination rule

```
B=.9; A=0.1;
[mSx mNx mUx]=bpafunction(DStestInputA, A,B);
[mSy mNy mUy]=bpafunction(DStestInputB, A,B);


function [mS mN mU]=bpafunction(cfv, A,B)
mS=B/(1-A)*cfv-A*B/(1-A);
mN=-B/(1-A)*cfv+B;

mS=mS.*[mS>=0];
mN=mN.*[mN>=0];
mU=1-mS-mN;


function [mSxy, mNxy, mUxy] =DSfusion(mSx, mNx, mUx, mSy, mNy, mUy)
% [mSx mNx mUx] mSx: nx1 vector

mSxy=(mSx.*mSy+mSy.*mUx+mSx.*mUy)./(1-mNx.*mSy-mSx.*mNy);
mNxy=(mNx.*mNy+mNx.*mUy+mUx.*mNy)./(1-mNx.*mSy-mSx.*mNy);

mUxy=(1-mSxy-mNxy);
```

A DS Case example with classification

```
cfv1org=1./(1+exp(-2.49*(v1-0.402))); % cfv for FEV1(v1)
cfvsorg=1./(1+exp(-2.48*(vs-0.403))); % cfv for symptom(vs)

% bpa for FEV1
B=0.9; A=0.1;
[mSx mNx mUx]=bpafunction(cfv1org, A,B);
[mSy mNy mUy]=bpafunction(cfvsorg, A,B);

%event window
    for i=1:lev
        evstart=(ev(i)-7*3);
        evend  =(ev(i)+7*1);
        if evstart<=0 evstart=1; end
        if evend > length(v) evend=length(v); end
        ew(evstart: evend)=1;

        plot((ev(i))*ones(1,2), [0; 1], 'r-', 'LineWidth',3)

    end
ew;
plot(1:length(v1), -ew/8, 'k--')

% DS2 fusion
[mSxy, mNxy, mUxy]=DSfusion(mSx, mNx, mUx, mSy, mNy, mUy);
EWmS=ew.*mSxy;
mSxy1=(mSxy-min(mSxy))/(1-min(mSxy)); %normalized to (0-1)
plot(1:length(v1), mSxy1, 'b', 'LineWidth',2)
```